

参赛密码 _____
(由组委会填写)

第十一届华为杯全国研究生数学建模竞赛

学 校	中南民族大学 华中师范大学 华中科技大学
--------	----------------------------

参赛队号	K0002
------	-------

队员姓名	1. 晏 磊
	2. 张 勇
	3. 叶子青

参赛密码 _____
(由组委会填写)



第十一届华为杯全国研究生数学建模竞赛

题 目 **D: 人体营养健康角度的中国果蔬发展战略研究**

摘 要:

本文综合考虑人体营养健康, 筛选确定了 8 类主要果蔬并预测了其消费量, 评价了我国居民目前营养素年摄入水平的合理度, 针对不同地区的居民提出了不同的最优购买策略, 并在考虑种植者收益和出口贸易的条件下确定了到 2020 年我国果蔬产品生产的调整战略。

数据预处理: 由于营养素摄入量属于居中型指标, 针对保障人体健康所需营养素参考值对果蔬中营养素含量进行无量纲居中型标准化处理; 根据参考文献对附件中错误数据进行更正, 对水果含量的单位和维生素 B12 的摄入这两类特殊情况进行说明。

问题 1: 首先基于营养素成分和含量, 分年龄段对常见果蔬进行聚类, 结果显示不同年龄段的聚类结果既有共性, 也有差异性, 能较好地体现人体营养素的需求共性和不用年龄段的需求差异性; 接下来基于组内营养素成分最大对果蔬进行等比例的筛选, 共筛选出主要水果 4 类: 苹果, 梨, 香蕉, 葡萄; 主要蔬菜 4 类: 菜花, 黄瓜, 香菇和胡萝卜; 然后根据 FAOSTAT 和 USDA 提供的数据, 确定了两种数据中消费量的计算公式, 针对 FAOSTAT, 消费量等于加工量与食物量之和; 针对 USDA, 消费量等于加工量和国内鲜果消费量之和, 结合果蔬损耗率, 确定了 1995-2011 年 8 类主要果蔬的消费量; 最后基于消费量的走势呈“指数型”增长的现象, 建立 GM(1,1)灰色预测模型对 2012-2020 年消费量进行预测, 残差检验表明预测结果精度较高。

问题 2: 对任意一种营养素, 若年均摄入量位于保障人体健康所需营养素参考范围值之内, 则定义我国居民该营养素的健康指数为 1, 若年摄入量偏离营养

素参考范围值越远,则定义我国居民该营养素的健康指数越小;进一步定义矿物质、维生素和膳食纤维的健康指数为其对应营养素的算术平均数,总营养素健康指数为矿物质、维生素和膳食纤维健康指数的算术平均数;然后将健康指数按照 1,[0.75,1],[0.5,0.75],[0.25,0.5],[0,0.25) 分为优秀、良好、合格、较差、非常差 5 个等级,以此构建模糊综合评价模型。结果显示自 1995-2011 年以来,我国居民矿物质、维生素和膳食纤维年摄入水平稳步上升;2012-2020 年,矿物质年摄入水平优秀,维生素年摄入水平良好,由于居民饮食趋于多样化,膳食纤维由于仅能通过蔬菜获取导致其健康指数在下降。

问题 3: 建立了分季度分地区的,以年购买成本最小为目标函数,各季度营养素摄入水平达到参考值为约束条件的单目标优化模型。首先假设果蔬的价格以年为周期稳定变化,以石家庄、威海和吴忠 3 市为例进行了模型的求解,结果显示:(1)不同地区的最优购买策略所需要的成本大体上相同,但具体金额会有所差别,这与当地的经济水平有关,(2)由于香菇在第二季度的价格远高于其他季度,因此第二季度的花费高于其他季度;为了考虑由于物价上涨等因素导致不同年份同一季度果蔬价格的变化,进一步引入居民价格消费指数 CPI,对果蔬价格进行修正,从而修正原优化模型。

问题 4: 建立了以居民购买成本最小、种植者利润最大和出口贸易量尽量大的为目标函数,以年均营养素摄入量达标、种植土地面积和进出口量为约束三目标优化模型。在模型求解中,建立线性回归模型预测了 2012-2020 年主要果蔬的年均产量,结合大量参考文献确定了其他模型参数,采用分层排序法对模型进行了求解,结果显示:截止到 2020 年,我国居民的主要果蔬年均摄入量已基本趋于稳定,说明问题 4 优化模型的宏观战略调控下,我国居民的人体营养均衡得到了较好的保证。

问题 5: 根据模型结果,从种植品种选择、引导居民消费量的购买、进出口贸易选择等方面对相关部门提供了 1000 字左右的政策建议。

最后,对模型进行了优缺点分析,并提出了改进方向。

关键词: 果蔬发展战略; 聚类分析模型; 综合评价模型; 优化模型; 线性回归; 分层排序法

1、问题重述

蛋白质、脂肪、维生素、矿物质、糖和水是人体需要的营养素。水果和蔬菜是重要的农产品，主要为人体提供矿物质、维生素、膳食纤维。近年来，中国水果和蔬菜种植面积和产量迅速增长，水果和蔬菜品种也日益丰富，中国居民生活水平不断提高，人们对人体营养均衡的意识也有所增强。然而多数中国居民喜食、饱食、偏食、忽视人体健康所需的营养均衡的传统饮食习惯尚未根本扭转，这就使得我国的果蔬消费（品种和数量）在满足居民身体健康所需均衡营养的意义下，近乎盲目无序，进而影响到果蔬生产。

预测我国果蔬的消费与生产趋势，科学地规划与调整我国果蔬的中长期的种植模式，具有重要的意义，在以某一种数据库为主，其余数据库为辅的基础上，需要考虑如下 4 个问题：

问题 1：在保证营养素成分和含量的基础上，结合附件数据和知名数据库，从众多果蔬品种中筛选主要的果蔬品种，估计其消费量并对消费量进行预测；

问题 2：根据我国居民矿物质、维生素、膳食纤维等营养素摄入现状，结合为保障人体健康所需要的各种营养成分的范围以及第一问的结论，评价中国居民目前矿物质、维生素、膳食纤维等营养的年摄入水平是否合理，并根据水果和蔬菜近期的消费趋势，预测至 2020 年中国居民的人体营养健康状况是趋于好转还是恶化？

问题 3：在保证营养均衡满足健康需要条件下，为当今中国居民（可以分区域分季度）提供主要的水果和蔬菜产品的按年度合理人均消费量，使得人们能够以较低的购买成本满足自身营养健康需求。

问题 4：为实现人体营养均衡满足健康需要，国家可能需要对水果和蔬菜各品种的生产规模做出战略性调整。既要考虑到居民人体的营养均衡，并使营养摄入量尽量在合理范围内；也要顾及居民的购买成本，使其购买成本尽量低；还要考虑种植者的收益；而且，作为国家宏观战略，还要考虑进出口贸易、土地面积等其他因素。建立数学模型重新计算中国居民主要的水果和蔬菜产品的按年度合理人均消费量，并给出到 2020 年我国水果和蔬菜产品生产的调整战略。

最后，请根据模型结果给相关部门提供 1000 字左右的政策建议。

2、模型假设

1. 除明显可发现错误外，附件数据真实可靠；
2. 本文以世界粮农组织数据库(FAOSTAT)数据库为主，其余数据库为辅；
3. 对于附件 4“中国居民膳食营养素参考摄入量”中涉及到的营养成分，若水果和蔬菜均不含该营养素，则认为该营养素应该由肉类补充，在本文建模中不予以考虑；
4. 对于维生素 B12，附件所提供果蔬品种中仅有李子含有，因此认为其他水果和蔬菜中不能为人体提供维生素 B12，若想从水果或蔬菜中摄取 B12，建议每日适当食用李子，在营养素摄入量指标中，不再考虑 B12；
5. 附件 EXCEL 中，若某种水果/蔬菜某种营养素成分含量为“-”，则认为该水果/蔬菜不含该营养成分。

3、符号说明

符号	含义
x_{ik}	第 i 种果蔬(每 100g)第 k 项营养素含量
$s_k(\delta)$	第 δ 类人群第 k 种营养素的日摄入量参考值
$\tilde{g}_i(k)$	考虑损耗率情况下第 i 种果蔬第 k 年的消费量(单位: 吨)
AM_k	我国居民第 k 种营养素的年均摄入量
R	营养素健康指数
$R_i(i=1,2,3)$	第 i 类营养素健康指数, $i=1,2,3$ 分别表示矿物质, 维生素和膳食纤维
$C_j(j=1,2,3,4)$	第 j 个季度购买果蔬的总花费(元)
CPI_{ij}	第 i 年相对于 2013 年第 j 个季度的 CPI

4、问题分析

本题模型的建立、求解与结果分析均应基于附件中的数据,因此在建模之前需要对数据进行预处理,主要包括数据标准化处理、错误数据的处理以及特殊情况说明等。在此基础上可进一步建模。

4.1 问题 1 的分析

本问要求从附件表格中筛选出主要的水果和蔬菜品种,并通过多种方法建立数学模型对其消费量进行估计,并研究其发展趋势。

考虑到备选水果和蔬菜数量众多,而有些蔬菜水果的营养素种类相近,食用功能相似,而且不同年龄段人群对营养素的要求不尽相同,因此首先针对不同年龄段人群(0-4; 5-18; 19-45; 46-),根据水果和蔬菜的营养素成分和含量对水果和蔬菜进行分别聚类;然后将每一类中水果和蔬菜中营养素成分和含量进行排序,按照一定的比例在每一类中确定若干代表,作为最终筛选出的水果和蔬菜,这样可以使得推荐筛选出来的水果针对性更强。

为了对消费量进行估计并进行消费发展趋势预测,首先参考 FAOSTAT(主要)和 USDA(辅助)中提供数据的表达形式以及果蔬资源的走向,根据他们提供的数据分析确定消费量的构成部分,综合考虑在采集、运输和生产等方面产生的损耗,

对消费量进行修正，即可确定中国居民主要果蔬的年均消费量。

为了预测消费量的发展趋势，有多种预测模型可供参考，然而不同的模型具有其适用范围，也需要进行预测结果检验。本问将判断主要果蔬消费量的发展趋势，根据发展趋势选择合适的数学模型对消费量进行预测。本问的思路流程图如图 1 所示。

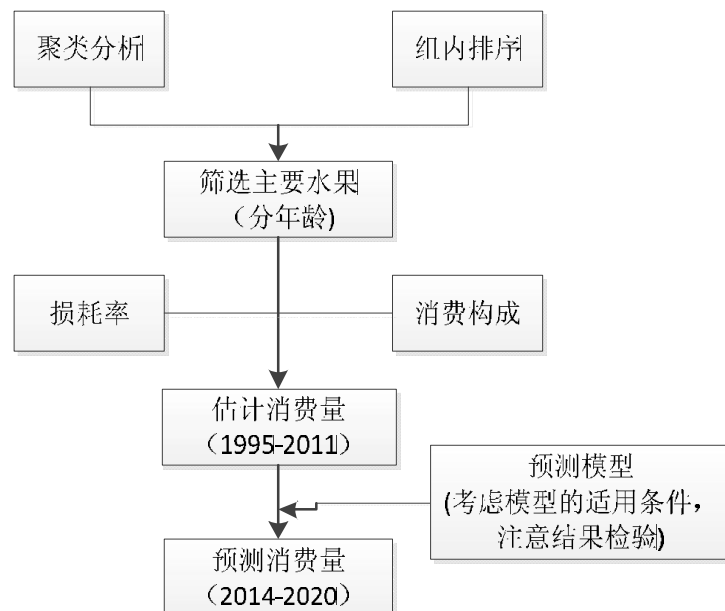


图 1: 问题 1 思路汇总图

4.2 问题 2 模型的建立与求解

本问要求根据问题 1 的结论，结合相关文献和附件数据，评价中国居民目前矿物质、维生素、膳食纤维等营养的年摄入水平是否合理，并根据水果/蔬菜的消费趋势，预测至 2020 年中国居民的人体营养健康状况是趋于好转还是恶化。

这是一个典型的评价模型。首先需要查阅为保障人体健康所需要各种营养成分的范围，其次需要根据附件数据及相关文献确定我国居民营养素的年摄入水平，最后根据摄入水平和参考范围之间的差距可以建立相应的评价模型。

由于为保障人体健康所需要各种营养成分是一个范围，而不是一个具体的数值，因此当摄入水平位于保证健康所需区间中时，可定义对应健康指数为 1，距离健康所需区间越远，健康指数越低，以此可建立模糊综合评价模型对我国居民目前矿物质、维生素、膳食纤维等营养的年摄入水平是否合理，并结合第 1 问的预测结果对 2020 年的发展趋势予以判断。

4.3 问题 3 模型的建立与求解

本问要求在保证营养均衡满足健康需要条件下，为当今中国居民提供合理的主要果蔬购买计划，使得购买成本最低。

这是一个典型的优化模型，考虑到不同地区、不同季度果蔬的价格可能存在差异性，因此分地区分季度进行购买政策的优化。目标函数即为总购买成本最小，约束条件为购买组合下摄入营养素达标，有此建立优化模型。

在本问中，价格数据的查找和变化趋势的确定是难点。为此，可以首先假设同一季度，同一水果不同年份的价格是稳定的，由此确定一种购买策略(称之为

静态购买策略); 另一方面, 考虑居民价格消费指数 CPI, 对购买价格进行修正, 可确定动态购买策略, 以供当地居民参考。

4.4 问题 4 模型的建立与求解

本问要求站在国家宏观战略角度上, 在尽量满足营养摄入量的前提下, 综合考虑居民购买能力种植者的利润以及出口贸易, 计算中国居民主要的水果和蔬菜产品的按年度合理人均消费量, 并给出到 2020 年我国水果和蔬菜产品生产的调整战略。

这也是一个典型的优化问题, 可建立以居民购买成本最小、种植者利润最大以及出口贸易尽量大为目标函数, 以营养摄入量达标和土地面积限制为约束条件的多目标优化模型。针对模型中的参数, 根据参考文献或历史数据的变化趋势预测其 2012-2020 年的取值, 从而对模型进行求解, 为我国宏观战略提供建议。

5、数据预处理

5.1 数据的标准化处理

附件 4 中提供的中国居民营养素参考日摄入量指标量纲差异性较大, 为了消除量纲差异性对结果带来的影响, 首先需要对数据进行标准化处理。

用 x_{ik} 表示第 i 种果蔬的第 k 项营养素的含量, 由于营养素摄入量应尽量在合理范围之内, 因此属于居中型指标。用 $s_k(\delta)$ 表示第 δ 类人群第 k 种营养素的日摄入量参考值, 则定义标准化后的第 i 种水果的第 k 项营养素的含量

$$x_{ik}^*(\delta) = \frac{s_k(\delta)}{|x_{ik} - s_k(\delta)|} \quad (6.1)$$

其中 $\delta=1,2,3,4$ 分别表示 4 种年龄段: 0-4 岁; 4-18 岁; 19-45 岁; 46 岁及以上。式(6.1)是一个无量纲化的数据, 其定义的合理性与可行性详见附录 A。为了书写和叙述上的方便性, 下文若不做特殊说明, 均用 $x_{ik}(\delta)$ 表示针对第 δ 类人群经过 (6.1) 式标准化处理后的第 i 种水果的第 k 项营养素的含量。

5.2 错误数据的处理

(1) 关于茄子的维生素 E 含量

附件 1 EXCEL 表格“常见蔬菜营养成分表”中, 100g 茄子的维生素 E 含量为 1013mg, 比其他蔬菜高出 3 个数量级, 明显不符合实际情况, 根据文献[1]的结论, 将其修改为 1.13mg。

(2) 关于 USDA 中的单位

USDA 中提供的水果蔬菜的单位为 MT(兆吨, 即 10^9kg), 而 FAOSTAT 提供的单位是 T(吨, 即 10^3kg), 且两个表格的数据在对应单位下的大小较为接近, 通过参考中华人民共和国农业部种植业管理司数据, 认为 USDA 中的数据单位有误, 在后文建模求解中, 默认 USDA 中的单位也为 T。

5.3 特殊情况说明

(1) 关于水果含量的单位

附件 1 中 EXCEL 表格“常见蔬菜营养成分表”中营养成分的含量均为每 100g 蔬菜的含量,而 EXCEL 表格“常见水果营养成分表”中营养成分未指明水果单位,通过查阅文献[1],并且对比“常见蔬菜营养成分表”,认为“常见水果营养成分表”中营养成分的含量也均为每 100g 水果的含量。

(2) 关于维生素 B12

附件中给出的所有水果和蔬菜的营养素成分及含量中,只有李子含有维生素 B12,因此认为其他水果和蔬菜中不能为人体提供维生素 B12,若想从水果或蔬菜中摄取 B12,建议每日适当食用李子。因此在营养素摄入量指标中,不再考虑 B12。

6、问题 1 模型的建立与求解

6.1 基于营养素成分和含量的聚类模型

6.1.1 聚类模型的建立

聚类分析是研究事物分类,将分类对象按照一定规则分成若干类的基本方法[2]。本节将基于营养素成分和含量对果蔬进行聚类。基于第五章中数据的处理,聚类分析的基本步骤如下[3]:

(1) 基于欧氏距离果蔬间相似性度量

在描述待分类的果蔬时,用第 i 种果蔬和第 j 种果蔬的欧氏距离 D_{ij} 来度量果蔬的相似程度,即

$$D_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2},$$

其中 x_{ik} 表示第 i 种果蔬的第 k 项营养素的含量,其中 $i=1,2,\dots,m$ 依次表示附件 1“可参考的营养成分表”文件夹下的 EXCEL 表格“常见水果的营养成分表”和“常见蔬菜的营养成分表”中 m 种果蔬, $k=1,2,\dots,n$ 表示附件 4 中 n 种营养素^①。

(2) 基于离差平方和法果蔬类间相似性度量

假设有两个已分好的果蔬类 G_1 和 G_2 ,同一类中不同果蔬之间的距离很小,它们就能很好地聚为一类,若两类的离差平方和足够大,则两类能够充分分离,因此可采用离差平方和对果蔬类间进行相似性度量。

设 G_1 和 G_2 类中分别有 m_1 和 m_2 种果蔬,由已知果蔬类中某种果蔬的 n 种营养素

^① 根据假设 3,本文仅考虑附件 4 中至少有一种水果和蔬菜含有的营养素

构成列向量：

$$g_s^1 = (g_{s1}^1, g_{s2}^1, \dots, g_{sn}^1)^T \in G_1,$$

$$g_r^2 = (g_{r1}^2, g_{r2}^2, \dots, g_{rm}^2)^T \in G_2,$$

$$g_t = (g_{t1}, g_{t2}, \dots, g_{tm})^T \in G_1 \cup G_2,$$

则果蔬各指标均值可构成列向量：

$$\bar{g}_s^1 = \left(\frac{1}{m_1} \sum_{s=1}^{m_1} g_{s1}^1, \dots, \frac{1}{m_1} \sum_{s=1}^{m_1} g_{sn}^1 \right)^T,$$

$$\bar{g}_r^2 = \left(\frac{1}{m_2} \sum_{r=1}^{m_2} g_{r1}^2, \dots, \frac{1}{m_2} \sum_{r=1}^{m_2} g_{rm}^2 \right)^T,$$

$$\bar{g}_t = \left(\frac{1}{m_1 + m_2} \sum_{t=1}^{m_1+m_2} g_{t1}, \dots, \frac{1}{m_1 + m_2} \sum_{t=1}^{m_1+m_2} g_{tm} \right)^T,$$

则 G_1 类中果蔬营养素的离差平方和

$$D_1 = \sum_{s=1}^{m_1} (g_s^1 - \bar{g}_s^1)^T (g_s^1 - \bar{g}_s^1),$$

G_2 类中果蔬营养素的离差平方和

$$D_2 = \sum_{r=1}^{m_2} (g_r^2 - \bar{g}_r^2)^T (g_r^2 - \bar{g}_r^2),$$

$G_1 \cup G_2$ 类中果蔬营养素的离差平方和

$$D_{12} = \sum_{t=1}^{m_1+m_2} (g_t - \bar{g}_t)^T (g_t - \bar{g}_t),$$

G_1, G_2 类间离差平方和

$$D(G_1, G_2) = D_{12} - D_1 - D_2,$$

若离差平方和 $D(G_1, G_2) = D_{12} - D_1 - D_2$ 越大，则 G_1, G_2 类分离得越充分，因此得到聚类分析模型如下：

$$\begin{cases} \max D(G_1, G_2) = D_{12} - D_1 - D_2 \\ D_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \end{cases}.$$

6.1.2 聚类模型的求解

水果共有 36 种，蔬菜共有 26 种，共考虑 4 种年龄段：0-4 岁；4-18 岁；19-45

岁；46 岁及以上。利用 SPSS 可以得到 4 种年龄段果蔬的共计 8 张聚类树状图，由于聚类图较大，正文仅给出 0-4 岁年龄段水果品种聚类图，其余 7 张图见附录 B 中，聚类结果如下表所示：

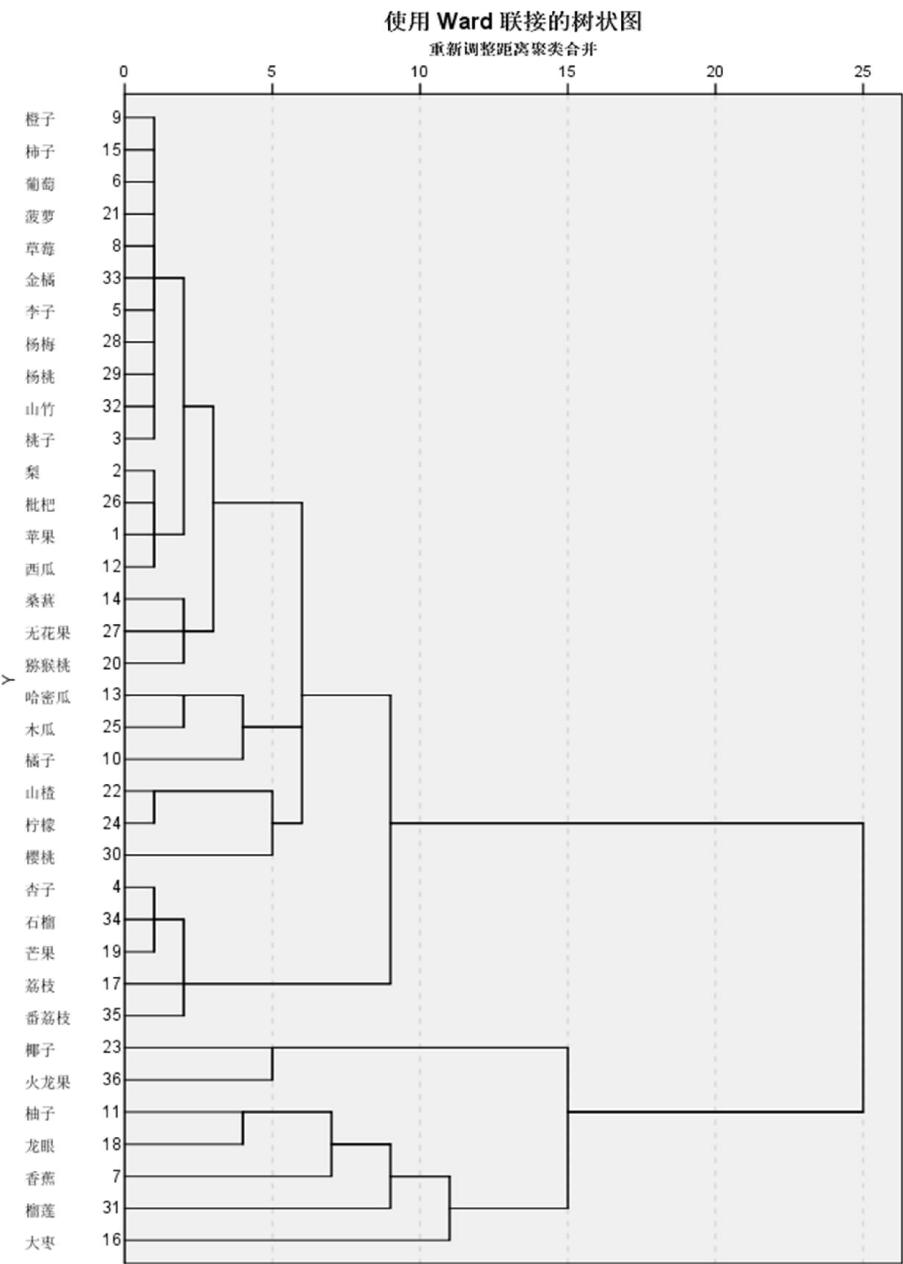


图 2：0-4 岁年龄段水果品种聚类图

表 1：0-4 岁年龄段所需果蔬聚类结果

	类别	所含品种
水果	第一类 (24 种)	樱桃, 柠檬, 山楂, 橘子, 木瓜, 哈密瓜, 猕猴桃, 无花果, 桑葚, 西瓜, 苹果, 枇杷, 梨, 桃子, 山竹, 杨桃, 杨梅, 李子, 金橘, 草莓, 菠萝, 葡萄, 柿子, 橙子
	第二类 (5 种)	番荔枝, 荔枝, 芒果, 石榴, 杏子
	第三类 (2 种)	火龙果, 椰子
	第四类 (5 种)	大枣, 榴莲, 香蕉, 龙眼, 柚子
蔬菜	第一类 (20 种)	苦瓜, 榨菜, 竹笋, 生菜, 韭菜, 油菜, 小白菜, 萝卜, 土豆, 南瓜, 黄瓜, 圆白菜, 芹菜, 大白菜, 蒜苗, 青椒, 西红柿, 丝瓜, 茄子, 冬瓜
	第二类 (1 种)	菠菜
	第三类 (2 种)	菜花, 胡萝卜
	第四类 (3 种)	香菇, 木耳, 蘑菇

表 2：5-18 岁年龄段所需果蔬聚类结果

	类别	所含品种
水果	第一类 (21 种)	山楂, 橘子, 木瓜, 猕猴桃, 无花果, 西瓜, 苹果, 枇杷, 梨, 杨桃, 杨梅, 李子, 金橘, 草莓, 菠萝, 葡萄, 柿子, 橙子, 芒果, 石榴, 龙眼
	第二类 (2 种)	火龙果, 椰子
	第三类 (3 种)	荔枝, 大枣, 榴莲
	第四类 (10 种)	樱桃, 柠檬, 哈密瓜, 桑葚, 山竹, 李子, 番荔枝, 杏子, 香蕉, 柚子
蔬菜	第一类 (17 种)	榨菜, 竹笋, 生菜, 韭菜, 萝卜, 土豆, 南瓜, 黄瓜, 圆白菜, 芹菜, 大白菜, 蒜苗, 西红柿, 丝瓜, 茄子, 冬瓜
	第二类 (3 种)	苦瓜, 青椒, 菜花
	第三类 (3 种)	油菜, 小白菜, 菠菜
	第四类 (3 种)	香菇, 木耳, 蘑菇

表 3: 19-45 岁年龄段所需果蔬聚类结果

	类别	所含品种
水果	第一类 (19 种)	橘子, 木瓜, 无花果, 西瓜, 苹果, 枇杷, 梨, 杨桃, 杨梅, 金橘, 草莓, 菠萝, 葡萄, 柿子, 橙子, 荔枝, 芒果, 石榴, 龙眼
	第二类 (1 种)	猕猴桃
	第三类 (13 种)	樱桃, 柠檬, 山楂, 哈密瓜, 桑葚, 桃子, 山竹, 李子, 荔枝, 杏子, 大枣, 香蕉, 柚子
	第四类 (3 种)	火龙果, 椰子, 榴莲,
蔬菜	第一类 (15 种)	苦瓜, 萝卜, 土豆, 南瓜, 黄瓜, 圆白菜, 芹菜, 蒜苗, 青椒, 西红柿, 丝瓜, 茄子, 冬瓜, 菜花, 胡萝卜
	第二类 (8 种)	榨菜, 竹笋, 生菜, 韭菜, 油菜, 小白菜, 大白菜, 菠菜
	第三类 (3 种)	香菇, 木耳, 蘑菇

表 4: 45 岁以上年龄段所需果蔬聚类结果

	类别	所含品种
水果	第一类 (13 种)	樱桃, 柠檬, 山楂, 哈密瓜, 桑葚, 桃子, 山竹, 杨桃, 杨梅, 李子, 荔枝, 大枣, 香蕉
	第二类 (2 种)	猕猴桃, 柚子
	第三类 (3 种)	火龙果, 椰子, 榴莲
	第四类 (18 种)	橘子, 木瓜, 无花果, 西瓜, 苹果, 枇杷, 梨, 桃子, 金橘, 草莓, 菠萝, 葡萄, 柿子, 橙子, 番荔枝, 芒果, 石榴, 龙眼
蔬菜	第一类 (15 种)	圆苦瓜, 萝卜, 土豆, 南瓜, 黄瓜, 圆白菜, 芹菜, 蒜苗, 青椒, 西红柿, 丝瓜, 茄子, 冬瓜, 菜花, 胡萝卜
	第二类 (8 种)	榨菜, 竹笋, 生菜, 韭菜, 油菜, 小白菜, 大白菜, 菠菜
	第三类 (3 种)	香菇, 木耳, 蘑菇

由

表 1-表 4 可知，不同年龄段的人群果蔬分类大致相同，也有一定的差别，例如香菇，木耳和蘑菇总是聚为一类，而荔枝和大枣在 0-4 岁人群中被分为 2 类，而在 5-18 岁人群中则聚为一类。聚类结果既考虑了不同年龄段推荐的果蔬既有共性，也有差异性，充分考虑到了不同年龄段需求上的共性和差异性，因此结果比较合理。

6.2 基于营养成分最大的主要水果/蔬菜的筛选

对于每一年龄段的人群，将常见水果和蔬菜均分为四大类，为了提取其中主要的水果，规定主要水果的种类应占该类水果的 20%，若某类水果的 20%尾数不足 1，则按照 1 个来计算。

设第 δ 类人群聚类结果中第 i 类有 $N_i(\delta)$ 种水果，则该类水果中应提取的主要水果数量为

$$\tilde{N}_i(\delta) = \begin{cases} 0.2N_i(\delta) & N_i(\delta) \equiv 0(\text{mod } 5) \\ \lceil 0.2N_i(\delta) \rceil + 1 & N_i(\delta) \not\equiv 0(\text{mod } 5) \end{cases},$$

其中 $[x]$ 为高斯函数，表示不超过 x 的最大正整数。

筛选出来的主要水果应该保证在本类中所占据的营养素含量最大，将各类结果所筛选出来的水果将其标准化后的营养素总含量从高到低进行排序，选取前 $\tilde{N}_i(\delta)$ 种水果，从而得到为了满足营养素的成分和含量，不同年龄段人群的主要水果与蔬菜结果，如下图所示：

表 5：不同年龄段人群推荐水果品种

年龄段	推荐水果品种
0-4 岁	樱桃，柠檬，橘子，木瓜，草莓，番荔枝，椰子，大枣
5-18 岁	苹果，梨，草莓，葡萄，火龙果，荔枝，哈密瓜，柠檬
19-45 岁	苹果，梨，葡萄，火龙果，荔枝，无花果，琵琶，猕猴桃
46 岁及以上	苹果，无花果，琵琶，猕猴桃，香蕉，哈密瓜，桑葚，金橘

表 6：不同年龄段人群推荐蔬菜品种

年龄段	推荐蔬菜品种
0-4 岁	苦瓜，南瓜，黄瓜，圆白菜，青椒，菠菜，菜花，蘑菇
5-18 岁	生菜，西红柿，黄瓜，圆白菜，菠菜，菜花，胡萝卜，蘑菇
19-45 岁	竹笋，芹菜，大白菜，蒜苗，丝瓜，竹笋，韭菜，菜花，香菇
46 岁及以上	竹笋，韭菜，土豆，蒜苗，丝瓜，大白菜，韭菜，菜花，香菇

- 由表 5 和表 6 可知，该分类：
- 涵盖了日常生活中常见的一些水果和蔬菜，诸如苹果，梨子，葡萄，香菇，菜花，黄瓜等；
 - 种类齐全，涵盖了温性水果和寒性水果；
- 不同年龄段推荐的水果/蔬菜既有共性，也有差异性，充分考虑到了不同年龄

段需求上的共性和差异性。主要体现在：

- 在推荐水果品种中，4 年龄段都推荐的有 0 种；3 种年龄段推荐的有 1 种(苹果)；2 种年龄段推荐的有 9 种，占总水果种类的 25%。
- 在推荐蔬菜品种中，4 种年龄段都推荐的有 1 种(菜花)；3 种年龄段都推荐的 0 种；2 种年龄段都推荐的有 10 种，占总蔬菜种类的 38.5%

因此上述结果能够较好地根据不同年龄段的人群提供果蔬进食建议。另一方面，由于本模型结果仅基于营养素成分和含量的聚类分析模型，未考虑到人群对水果的具体需求与偏好，显然结果存在一定的局限性，例如 0-4 岁婴幼儿推荐蔬菜品种包括苦瓜，而对于该年龄段的婴幼儿而言，是不喜欢苦瓜的口味的，此时则建议家长使用榨汁机进行处理。

为了整体考虑果蔬发展战略研究的需要，不区分年龄段的差异性，对果蔬进行基于营养素成分和含量的聚类，综合考虑市面上的普及程度，得到整体推荐主要果蔬品种结果，如下表所示。

表 7：主要推荐果蔬品种

类别	主要推荐品种
水果	苹果，梨，葡萄，香蕉，
蔬菜	黄瓜，菜花，胡萝卜，香菇，

表 7 推荐的果蔬在营养素含量上覆盖了其他所有果蔬含有的矿物质、维生素和膳食纤维，且口味较为大众化，具有一定的代表性。

6.3 消费量的估计

为了保持中国居民生产、消费等数据的一致性，根据各数据库数据的完整性与时间跨度的久远性，本文主要使用世界粮农组织发布的数据(FAOSTAT)。在 4 类主要水果(苹果，梨，香蕉，葡萄)中，FAOSTAT 提供了苹果，香蕉和葡萄的数据；而 USDA 提供了梨的数据，为了对消费量进行估计，首先整合两个数据库提供的不同表达形式的指标，再通过分析确定居民消费量。

FAOSTAT 提供数据年份跨度从 1961 年-2011 年，USDA 提供梨的数据年份跨度从 1995-2013 年，为了保持数据前后维度的一致性，本文选取 1995-2011 年的数据进行消费量的估计。

6.3.1 消费量的定义及说明

(1) 世界粮农组织数据库数据指标分析

世界粮农组织数据库(FAOSTAT)提供了果蔬的 8 项指标，包括：Production(生产量)，Import Quantity(进口量)，Stock Variation(库存变化)，Export Quantity(出口数量)，Domestic Supply Quantity(国内供应量)，Waste(浪费)，Processing(加工)，Food(食品)，以及 Other Util(其他用途)。

通过整理 FAOSTAT 中的相关数据，可知上述 8 项指标具有如下关系：

- 生产量+进口量=国内供应量+出口量；
- 国内供应量=浪费量+加工量+食物+其他用途。

结合文献[6-8]的结论，浪费量和其他用途主要指未能被居民利用的产量，而食物和加工量分别以直接和间接的方式来表示消费量，其中食物属于直接消费量，例如进食苹果、西红柿等水果和蔬菜；而加工量则以间接的方式表示消费量，例

如苹果通过加工可以成为苹果醋，通过饮用苹果醋即间接地消费了苹果。因此认为国内居民消费量包括 2 部分：食物和加工量。上述分析示意图见图 3。

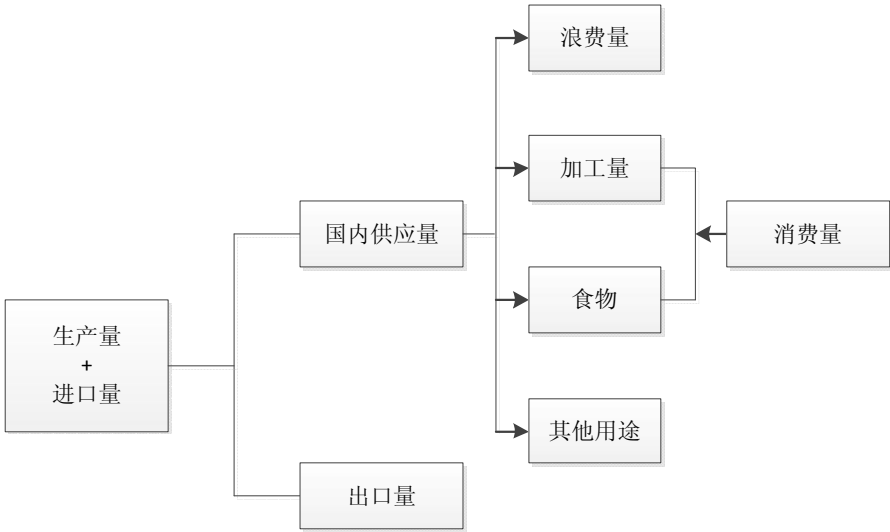


图 3：水果/蔬菜消费量组成示意图(来源：FAOSTAT)

(2) 美国农业部数据库数据指标分析

美国农业部(USDA)提供了水果，蔬菜的 9 项指标，包括：Total Distribution(总分布量)，For Processing(加工量)，Exports(出口量)，Fresh Dom. Consumption(国内鲜果消费量)，Total Supply(总供应量)，Imports(进口量)，Production(生产量)，Non-Comm. Production(非商业生产量)，Commercial Production(商业生产量)

通过整理 FAOSTAT 中的相关数据，可知上述 9 项指标具有如下关系：

- 总分布量=总供应量=进口量+生产量=出口量+加工量+国内鲜果消费量
- 生产量=商业生产量+非商业生产量

分析理由同上，认为国内鲜果消费量与加工量即为国内居民消费量，指标关系如下图所示：

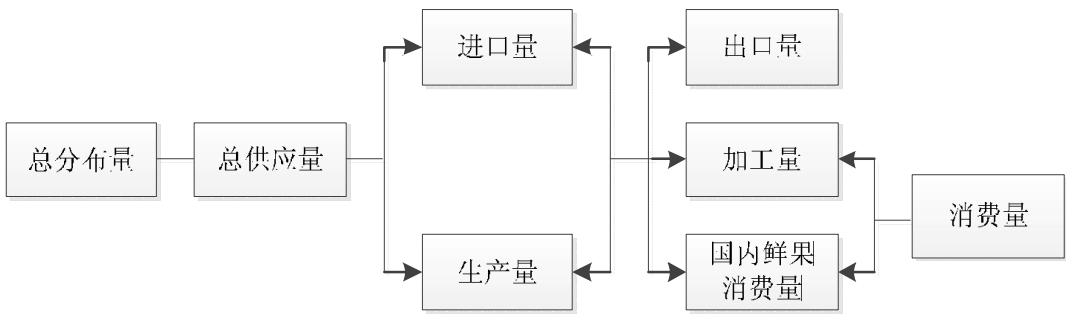


图 4：水果/蔬菜消费量组成示意图(来源：USDA)

(3) 消费量的定义

针对 FAOSTAT 所能提供的果蔬数据，定义该果蔬的消费量为加工量与食物量之和。

针对 FAOSTAT 未能提供单 USDA 所能提供的果蔬数据，定义该果蔬的消费量为加工量与国内鲜果消费量之和。

6.3.2 基于用途走向和损耗率的消费量估计

根据 FAOSTAT 计算苹果，葡萄，和香蕉的消费量，根据 USDA 计算梨的消

费量，从而得到表 8。

表 8：1995-2011 年中国居民主要果蔬消费量估计-未考虑损耗率(单位：吨)

年份	苹果	梨	葡萄	香蕉	菜花	黄瓜	香菇	胡萝卜
1995	12470493	4858392	1894052	3110626	8892146	28429748	13923292	23239550
1996	15352114	3866400	2037155	2847152	10032324	28436581	16042350	23399150
1997	15329710	6318600	2152325	3243133	7983454	38721847	18162387	27340800
1998	17150761	7167364	2440786	3810254	8803733	43238957	18452867	31026750
1999	18613671	7643077	2856812	4346274	8523790	40964632	17696281	32360550
2000	17644167	8233093	3404887	5163170	9021437	47277674	23176121	35489300
2001	17627788	8602300	3768160	5321392	9292744	49761797	26291969	34495500
2002	16671474	9037700	4564012	5540051	9764114	53246726	27677054	37563200
2003	17303246	9495700	5216970	5903111	10704929	53154505	32386540	32835550
2004	20068750	10300100	5700082	5948824	11140328	58864211	33924659	33414650
2005	20116235	10961200	5793382	6338803	11270479	58472762	38321642	35216250
2006	21789829	11588100	6228647	6631559	11455390	45591405	40756076	43976050
2007	22024105	12471700	6593620	7156319	13610150	53781859	42157593	37039400
2008	22655268	13092500	7017067	6791451	14676517	58517549	44978248	34849600
2009	24798955	13793000	7742276	7656661	14745291	59978464	49757321	40365550
2010	25968754	14634300	8365835	8482011	15472227	65074158	56908143	47631100
2011	28975897	15382700	8886724	9305732	17050116	66101928	61606237	48189150

6.3.3 考虑损耗率的消费量估计

图 3 和图 4 给出的消费量组成示意图能够较好地反映消费量的构成以及进出口水果的流向等，然而未考虑到水果的损耗率。根据附件 3 的数据，各种果蔬从田间地头到大地市场，从大地市场到零售市场，从零售市场到餐桌均具有消耗率，用 α_{ij} 表示第 i 种果蔬第 j 种过程的损耗率。为了准确估计消费量，应对 6.3.1 中估计的消费量进行修正。

从田间地头到大地市场的水果，既有最终用作消费量的部分，也有浪费和用作其他用途的部分，因此损耗率 α_{i1} 一部分用于消费量的损耗，另一部分用于浪费和其他用途的损耗。

根据 FAOSTAT 提供的数据，计算得到 1995-2011 年加工量和食物量占出口和国内供应量的百分比如下：

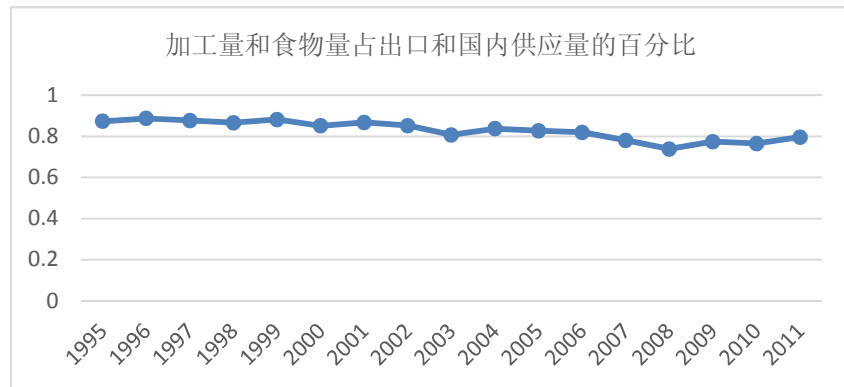


图 5：加工量和食物量占出口和国内供应量总和的百分比（数据来源：FAOSTAT）

从上图可以看出，加工量和食物量占出口和国内供应量的百分比基本稳定，均值为 83%，因此假设损耗率 α_1 的 83% 用于消费量的损耗，而剩余的 17% 用于浪费和其他用途的损耗；从大地市场到零售市场，从零售市场到餐桌的消耗率均规定为消费量的损耗。

则基于损耗率修正后的第 i 种水果第 k 年的消费量

$$\tilde{g}_i(k) = \frac{g_i(k)}{(1 - 0.83\alpha_1)(1 - \alpha_2)(1 - \alpha_3)},$$

计算得到 1995-2011 年中国居民主要水果消费量估计，如表 9 所示：

表 9：1995-2011 年中国居民主要果蔬消费量估计-考虑损耗率(单位：吨)

年份	苹果	梨	葡萄	香蕉	菜花	黄瓜	香菇	胡萝卜
1995	13044591	5526068	2897203	4692435	9823378	31407059	15381412	32613397
1996	16058872	439775	3116098	4294980	11082961	31414608	17722388	32837373
1997	16035437	7186949	3292265	4892325	8819523	42777001	20064447	38368917
1998	17940323	8152356	3733504	5747837	9725706	47767166	20385347	43541622
1999	19470580	8693446	4369871	6556433	9416446	45254662	19549528	45413421
2000	18456444	9364546	5208224	7788735	9966209	52228838	25603245	49804175
2001	18439311	9784492	5763898	8027416	10265929	54973112	29045401	48409519
2002	17438971	10279728	6981259	8357267	10786663	58823001	30575539	52714599
2003	18099828	10800670	7980044	8904950	11826005	58721122	35778227	46080016
2004	20992646	11715616	8719028	8973908	12307001	65028778	37477426	46892700
2005	21042317	12467570	8861743	9562199	12450782	64596334	42334884	49420989
2006	22792958	13180623	9527538	10003827	12655058	50365974	45024265	61714120
2007	23038019	14185654	10085812	10795436	15035476	59414175	46572556	51979520
2008	23698238	14891769	10733530	10245026	16213518	64645812	49688604	48906448
2009	25940613	15688537	11842833	11550211	16289495	66259721	54968167	56647298
2010	27164266	16645454	12796649	12795267	17092559	71889063	62867860	66843462
2011	30309847	17496705	13593418	14037865	18835693	73024466	68057963	67626605
年均增长率	5.67%	8.45%	10.27%	7.34%	4.51%	6.09%	10.00%	5.24%

从表 9 看出,近年来,我国居民的主要果蔬消费量逐年递增,且增幅较为明显,在 8 类果蔬中,葡萄和香菇的年均增长率均超过 10%,菜花的年均增长率最低,为 4.51%,8 种果蔬的平均年增长率达到 7.2%。

6.4 消费量发展趋势预测——基于 GM(1,1)模型

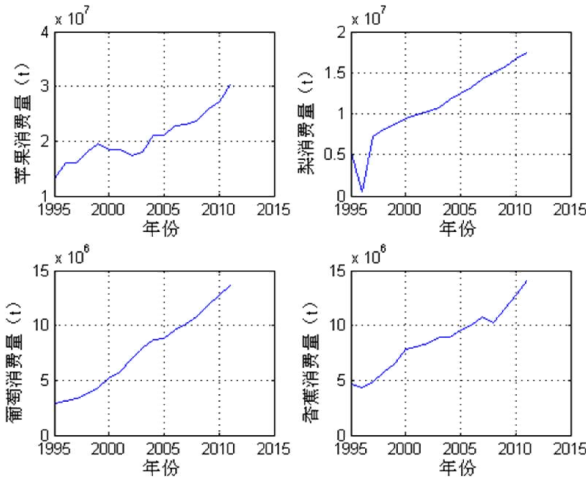


图 6: 1995-2011 年主要水果消费量

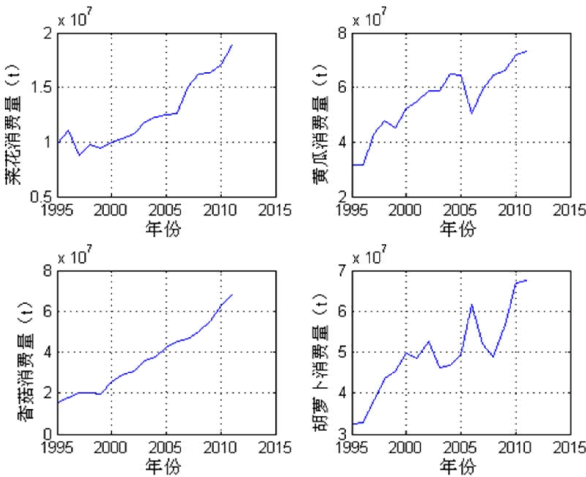


图 7: 1995-2011 年主要蔬菜消费量

图 6 和图 7 给出了自 1995-2011 年来,我国主要果蔬的年消费量,由图可知,主要水果消费量逐年递增,且近似呈“指数式增长”,因此本文可使用 GM(1,1)模型进行消费量发展趋势的预测^[4]。

6.4.1 数据的检验与处理

为了保证建模方法的可行性,需要对已知数据列进行必要的检验处理。设参考数据列(即 1995-2011 年的主要水果消费量)为 $g^{(0)} = (g^{(0)}(1), g^{(0)}(2), \dots, g^{(0)}(n))$, 计算数列的级比

$$\lambda(k) = \frac{g^{(0)}(k-1)}{g^{(0)}(k)} \quad (k=2,3,\dots,n)$$

如果所有的级比 $\lambda(k) \in G = (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$ 内, 则可将数据进行灰色预测, 否则, 需要取适当的常数 c , 令

$$y^{(0)}(k) = g^{(0)}(k) + c \quad (k=1,2,\dots,n)$$

使得新数据列的级比 $\lambda'(k) \in G = (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$ 。

6.4.2 GM(1,1)模型的建立

已知参考数列为 $g^{(0)} = (g^{(0)}(1), g^{(0)}(2), \dots, g^{(0)}(n))$, 做一次累加 (AGO) 生成数列

$$\begin{aligned} g^{(1)} &= (g^{(1)}(1), g^{(1)}(2), \dots, g^{(1)}(n)) \\ &= (g^{(1)}(1), g^{(1)}(1) + g^{(0)}(2), \dots, g^{(1)}(n-1) + g^{(0)}(n)) \end{aligned}$$

其中 $g^{(0)}(1) = g^{(1)}(1)$, $g^{(1)}(k) = \sum_{i=1}^k g^{(0)}(i) \quad (k=1,2,\dots,n)$ 。

求均值 (mean) 数列

$$z^{(1)}(k) = 0.5g^{(1)}(k) + 0.5g^{(1)}(k-1) \quad (k=2,3,\dots,n)$$

即 $z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$ 。于是建立灰微分方程为

$$g^{(0)}(k) + az^{(1)}(k) = b \quad (k=2,3,\dots,n)$$

记 $u = (a, b)^T$, $Y = (g^{(0)}(2), g^{(0)}(3), \dots, g^{(0)}(n))^T$

$$B = \begin{pmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{pmatrix}$$

由最小二乘法拟合得到 $\begin{pmatrix} a \\ b \end{pmatrix} = (B^T B)^{-1} B^T Y$, 进而解得

$$g^{(1)}(k+1) = (g^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \quad (k=1,2,\dots,n-1)$$

所以预测值

$$\hat{g}^{(1)}(k+1) = (g^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \quad (k=1,2,\dots,n-1)$$

而且

$$\hat{g}^{(0)}(k+1) = \hat{g}^{(1)}(k+1) - \hat{g}^{(1)}(k) \quad (k=1,2,\dots,n-1)$$

6.4.3 GM(1,1)模型的求解与检验

(1) 模型的求解

将表 9 的数据代入到 GM(1,1)模型中(MATLAB 源程序代码见附录 C)，得到 2012-2020 年水果消费量的预测结果，如表 10 所示。

表 10: 2012-2020 年主要果蔬消费量预测(单位: 吨)

年份	苹果	梨	葡萄	香蕉	菜花	黄瓜	香菇	胡萝卜
2012	29078847	18920113	15742247	14537923	18845998	76011691	74116240	66164360
2013	30271654	20112948	17195716	15467054	19817881	78793758	80958330	68403681
2014	31513390	21380986	18783383	16455567	20839884	81677650	88432051	70718791
2015	32806061	22728968	20517637	17507256	21914592	84667094	96595715	73112256
2016	34151758	24161935	22412014	18626159	23044721	87765954	105513012	75586727
2017	35552654	25685244	24481297	19816572	24233132	90978233	115253516	78144946
2018	37011015	27304592	26741636	21083066	25482828	94308083	125893221	80789748
2019	38529198	29026033	29210669	22430503	26796971	97759807	137515137	83524063
2020	40109656	30856003	31907667	23864055	28178884	101337867	150209938	86350920

其中梨的预测曲线与原始值的拟合效果如图 8 所示，从图 8 可以看出，拟合效果是非常好的。其余果蔬的预测曲线与原始值的拟合效果见附录 D。

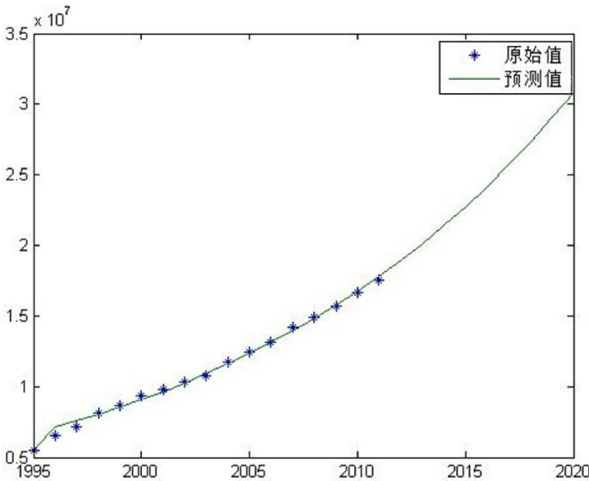


图 8: 梨的预测值与原始值对比曲线图(横坐标表示年份, 纵坐标表示消费量, 单位: 吨)

(2) 残差检验

为了检验预测结果是否合理，需要对模拟值与实际值做残差检验，其中 $g^0(k)$ 表示第 k 年的实际数值， $\hat{g}^0(k)$ 表示第 k 年的预测值，则残差

$$\varepsilon(k) = g^0(k) - \hat{g}^0(k),$$

相对误差

$$\Delta(k)=\frac{|\varepsilon(k)|}{g^0(k)}.$$

当相对误差 $\Delta(k)<0.2$ 时，认为结果达到要求，当 $\Delta(k)<0.1$ 时，认为结果具有较高的精度^[4]。

对水果和蔬菜的结果做残差分析，结果如下表所示。

表 11：残差检验表(水果)

年份	苹果	梨	葡萄	香蕉	菜花	黄瓜	香菇	胡萝卜
1996	0.05	0.10	0.07	0.19	0.15	0.17	0.02	0.18
1997	0.01	0.05	0.18	0.17	0.01	0.04	0.02	0.05
1998	0.08	0.01	0.12	0.06	0.04	0.04	0.06	0.05
1999	0.11	0.02	0.14	0.01	0.04	0.05	0.15	0.05
2000	0.03	0.03	0.05	0.11	0.03	0.05	0.00	0.11
2001	0.01	0.01	0.03	0.08	0.06	0.07	0.03	0.05
2002	0.12	0.00	0.07	0.06	0.06	0.10	0.00	0.10
2003	0.12	0.01	0.11	0.07	0.01	0.06	0.06	0.06
2004	0.00	0.01	0.11	0.01	0.02	0.12	0.02	0.08
2005	0.04	0.01	0.04	0.01	0.06	0.09	0.06	0.06
2006	0.00	0.01	0.03	0.00	0.10	0.22	0.03	0.12
2007	0.03	0.02	0.00	0.01	0.03	0.07	0.02	0.08
2008	0.04	0.01	0.03	0.11	0.05	0.02	0.05	0.18
2009	0.01	0.00	0.02	0.05	0.01	0.03	0.03	0.06
2010	0.01	0.01	0.03	0.00	0.00	0.02	0.01	0.07
2011	0.08	0.02	0.06	0.03	0.05	0.00	0.00	0.05

表 11 中仅有黄瓜 2006 年的预测值残差略微超过了 0.2，为 0.22，其余的残差均小于 0.2，由于数据庞大，且涉及到不同数据库之间数据的使用，因此可能产生一定的误差，故可认为模型精度达到要求。

从表 9-表 10 的结果可以看出：

- 近年来，在随着我国经济的发展和水果蔬菜种植面积和产量迅速增长、背景下，果蔬的消费量稳步提高，反映了居民生活水平的不断提高；
- 作为一个农业大国，目前我国蔬菜的消费量要明显高于水果消费量。

7、问题 2 模型的建立与求解

7.1 综合评价模型的建立

用 $S_k=[S_{k\min},S_{k\max}]$ 表示为保障人体健康所需要的第 k 种营养素的区间， x_{ik} 表示第 i 种果蔬第 k 种营养素的含量， M_i 表示我国居民第 i 种果蔬的年均摄入量(单位：

100g/人), 则我国居民第 k 种营养素的年均摄入量

$$AM_k = \sum_{i=1}^n M_i x_{ik}, \quad (7.1)$$

其中 $M_i = \frac{10^4 \tilde{g}_i}{N}$, \tilde{g}_i 为第 i 种果蔬的消费量 (单位: t), N 为所考虑年份的总人数

(单位: 人), 系数 10^4 是为了将 M_i 的单位化为 100g/人。

当 $AM_k \in S_k$ 时, 认为已经达到健康标准, 定义第 k 种营养素的健康指数 $r_k = 1$;

当 $AM_k \notin S_k$ 时, 有可能年均摄入量偏大, 也有可能年均摄入量偏小, 均属于非健康状态, 且当 AM_k 与 S_k 偏离得越远的时候, 健康指数越低。

更一般地, 定义第 k 种营养素的健康指数

$$r_k = \begin{cases} 1 & AM_k \in S_k \\ \frac{AM_k}{S_{k \min}} & AM_k < S_{k \min} \\ \frac{S_{k \max}}{AM_k} & AM_k > S_{k \max} \end{cases}, \quad (7.2)$$

显然 $r_k \leq 1$ 是一个无量纲指标。

营养素共分为 3 大类: 矿物质(9 类), 维生素(10 类)和膳食纤维(1 类)^②, 分别对应 $k=1, \dots, 9; k=10, \dots, 19$ 和 $k=20$ 。每一类营养素的健康指数定义为该类各营养素健康指数的平均值, 即

$$\text{矿物质健康指数: } R_1 = \frac{1}{9} \sum_{k=1}^9 r_k;$$

$$\text{维生素健康指数: } R_2 = \frac{1}{10} \sum_{k=10}^{19} r_k;$$

$$\text{膳食纤维健康指数: } R_3 = r_{20}。$$

进一步定义营养素健康指数

$$R = \frac{1}{3}(R_1 + R_2 + R_3) \quad (7.3)$$

显然 R 及 R_i 均为无量纲指标且 $R \leq 1$ 恒成立, 根据模糊数学隶属度的思想, 将健康指数划分为“优秀”, “良好”, “合格”, “较差”, “非常差”五个等级。

综上所述, 得到我国居民营养素的年均摄入量的合理性评价模型:

^②根据模型假设 3, 此处仅考虑蔬菜或水果含有的营养素, 附件 4 中提供但是果蔬不含有的营养素认为应该从肉类食物或粮食中获取。

$$\begin{cases} R=1 & \text{优秀} \\ R \in [0.75, 1) & \text{良好} \\ R \in [0.5, 0.75) & \text{合格} \\ R \in [0.25, 0.5) & \text{较差} \\ R \in [0, 0.25) & \text{非常差} \end{cases}, \quad (7.4)$$

其中

$$\begin{cases} R = \frac{1}{3}(R_1 + R_2 + R_3); \quad R_1 = \frac{1}{9} \sum_{k=1}^9 r_k; \quad R_2 = \frac{1}{10} \sum_{k=10}^{19} r_k; \quad R_3 = r_{20} \\ r_k = \begin{cases} 1 & AM_k \in S_k \\ \frac{AM_k}{S_{k \min}} & AM_k < S_{k \min} \\ \frac{S_{k \max}}{AM_k} & AM_k > S_{k \max} \end{cases} \\ AM_k = \sum_{i=1}^n M_i x_{ik}; \quad M_i = \frac{10^4 \tilde{g}_i}{N}; \quad S_k = [S_{k \min}, S_{k \max}] \end{cases}. \quad (7.5)$$

7.2 参数的确定

7.2.1 保障人体健康所需要的营养素 s_k 的确定

附件 4 提供了不同年龄段中国居民膳食营养素参考日摄入量，定义保证人体健康所需要的营养素 $s_{k \max}$ 和 $s_{k \min}$ 为第 k 类营养素在不同年龄段的参考日摄入量的上界和下界，从而得到保障人体健康所需要的矿物质，维生素和膳食纤维^[9]的范围，如下表所示。

表 12：保障人体健康所必须的每日营养素含量范围

矿物质	钙(mg)	磷(mg)	钾(mg)	钠(mg)	镁(mg)	铁(mg)	锌(mg)	硒(ug)	铜(mg)	——
	[300,1200]	[150,1000]	[500,2500]	[200,2200]	[30,400]	[0.3,25]	[1.5,21.5]	[15,65]	[0.4,2]	——
维生素	A(mg)	C(mg)	E(mg)	B1(ug)	B2(ug)	B6(mg)	泛酸(mg)	叶酸(ug)	烟酸(ug)	生物素(ug)
	[400,1200]	[40,130]	[3,14]	[0.2,1.8]	[0.5,1.7]	[0.1,1.9]	[1.7,7]	[65,600]	[2.0,18]	[5,35]
膳食纤维	25-35g									

表 12 给出了保障人体健康所必须的每日矿物质含量范围，将各数值乘以 365 即可得到保障人体健康所必须的年均营养素含量范围。

7.2.2 居民年摄入水平的确定

FAOSTAT 和 USDA 提供了各类水果的消费量，而在问题 1 中筛选出了 8 类主要果蔬，为了通过 8 类主要果蔬反映整体的居民年均摄入水平，需要对数据进行等比例扩张。

同问题 1，考虑 1995-2011 年的数据，用 G_i 表示第 i 年由 FAOSTAT 和 USDA

提供的水果种类并集的消费量，在 6.3.3 中已经确定了第 i 年第 j ($j=1,2,3,4$) 种水果的消费量 g_{ij} ，则定义等比例扩张后第 i 年第 j ($j=1,2,3,4$) 种水果的消费量

$$\tilde{g}_{ij} = \frac{G_i}{\sum_{j=1}^4 g_{ij}} \times g_{ij}, \quad (7.6)$$

则第 i 年第 k 种营养素的年均摄入量

$$AM_{ik} = \sum_{j=1}^4 (\tilde{g}_{ij} \cdot x_{jk}), \quad (7.7)$$

其中 x_{jk} 表示第 j 种水果第 k 种营养素。

7.2.3 人口 N 的确定

根据中华人民共和国统计局数据^[11]，做出 1995-2013 年我国人口变化示意图，如图 9 所示。

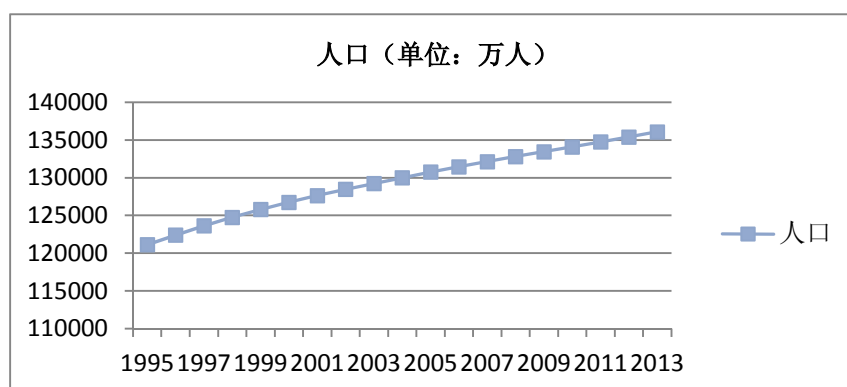


图 9：我国 1995-2013 年人口变化示意图

对于 1995-2013 年的人口数据，可以直接通过查表得到，为了对到 2020 年我国居民健康状况进行评价，需要预测 2014-2020 年的我国人口。

根据图 9 的人口走势图，可认为人口变化几乎呈线性增长。利用 MATLAB 进行最小二乘拟合得到我国第 i 年人口预测值

$$N_i = 800i - 1.47 \times 10^6,$$

其中 $i=1995, 1996, \dots, 2020$ 分别表示年份。

拟合检验的结果显示 $R^2 = 0.9876$ ，调整 $\tilde{R}^2 = 0.9869$ 接近于 1，拟合优度 RMSE=518.1，因此该线性拟合效果较好，拟合结果通过检验。

因此 2014-2020 年我国人口的预测值为：

表 13：2014-2020 年我国人口预测值(单位：万人)

年份	2014	2015	2016	2017	2018	2019	2020
人口	141200	142000	142800	143600	144400	145200	146000

7.3 我国居民年均营养素健康指数的确定及 2020 年营养健康状况预测

7.3.1 我国居民年均营养素健康指数的求解

代入 1995-2011 年相关数据，结合(7.1)-(7.7)，计算得到 1995-2011 年我国居民年均营养素健康指数如下表所示

表 14：1995-2011 年我国居民年均营养素健康指数汇总表

	营养素		矿物质		维生素		膳食纤维	
	指数	等级	指数	等级	指数	等级	指数	等级
1995	0.65	合格	0.82	良好	0.82	良好	0.30	较差
1996	0.68	合格	0.84	良好	0.84	良好	0.34	较差
1997	0.70	合格	0.86	良好	0.86	良好	0.39	较差
1998	0.73	合格	0.88	良好	0.89	良好	0.41	较差
1999	0.74	合格	0.90	良好	0.92	良好	0.40	较差
2000	0.79	良好	0.92	良好	0.94	良好	0.51	合格
2001	0.82	良好	0.93	良好	0.95	良好	0.58	合格
2002	0.84	良好	0.94	良好	0.95	良好	0.61	合格
2003	0.87	良好	0.96	良好	0.96	良好	0.71	合格
2004	0.89	良好	0.98	良好	0.95	良好	0.75	合格
2005	0.93	良好	1.00	良好	0.95	良好	0.84	良好
2006	0.95	良好	1.00	良好	0.94	良好	0.90	良好
2007	0.96	良好	1.00	优秀	0.95	良好	0.93	良好
2008	0.98	良好	1.00	优秀	0.95	良好	1.00	良好
2009	0.98	良好	0.99	良好	0.94	良好	1.00	优秀
2010	0.97	良好	0.98	良好	0.92	良好	1.00	优秀
2011	0.96	良好	0.97	良好	0.91	良好	1.00	优秀

从表 14 看出，截止到 2011 年，随着人们对人体营养均衡意识的加强，我国居民的矿物质、维生素和膳食纤维的摄入量在稳步靠近于参考标准，居民整体营养素健康指数也从 0.68（合格）上升至 0.98(良好)，并有极微弱的反弹波动。

7.3.2 2020 年营养健康状况预测

在 6.4 中，基于 GM(1,1)模型预测了 2012-2020 年的果蔬消费量，结合(7.1)-(7.7)，预测得到 2012-2020 年我国居民年均营养素健康指数如下表所示：

表 15：2012-2020 年我国居民营养健康状况预测

	营养素		矿物质		维生素		膳食纤维	
	指数	等级	指数	等级	指数	等级	指数	等级
2012	0.98	良好	1.00	优秀	0.94	良好	1.00	优秀
2013	0.97	良好	1.00	优秀	0.93	良好	0.98	良好
2014	0.93	良好	1.00	优秀	0.90	良好	0.89	良好
2015	0.90	良好	1.00	优秀	0.88	良好	0.82	良好
2016	0.87	良好	1.00	优秀	0.86	良好	0.76	良好

2017	0.85	良好	1.00	优秀	0.84	良好	0.70	合格
2018	0.82	良好	1.00	优秀	0.82	良好	0.64	合格
2019	0.80	良好	1.00	优秀	0.80	良好	0.59	合格
2020	0.78	良好	1.00	优秀	0.79	良好	0.55	合格

由表 15 可以看出：

- 我国居民营养素年均摄入水平在整体上为良好，略微偏离参考范围
- 我国居民矿物质的年摄入水平(优秀)已达到参考范围；
- 我国居民维生素的年摄入水平(良好)略微偏离于参考范围；
- 我国居民膳食纤维的年摄入水平在 2016 年之前良好，在 2017 年及以后开始下降，原因之一在于膳食纤维仅能通过蔬菜提供，缺乏多样性。
- 此外，注意到无论是营养素，还是维生素和膳食纤维，其健康指数在逐步下降。这是因为在消费量的预测中采用的是呈“指数型”增长的 GM(1, 1)模型，预测时间越远，预测消费量数值越大(呈指数增长)，而人口增长的速度远远不及消费量的增长速度，使得营养素摄入过剩，由(7.2)知这样会降低健康指数取值，因此本模型只适用于短期预测，对于长期预测，误差相对较大。

8、问题 3 模型的建立与求解

8.1 基于成本最小的主要果蔬购买策略的优化模型

8.1.1 目标函数的确定

一年 4 个季度，用 p_{ij} 表示所考虑区域第 i 种果蔬第 j 个季度的平均价格(元/kg)，

其中 $i=1,2,\dots,8$ 分别表示苹果，梨，葡萄，香蕉，菜花，黄瓜，香菇和胡萝卜，

设第 j 个季度应购买第 i 种果蔬的数量为 m_{ij} (kg)，则第 j 个季度购买果蔬的花费

$$C_j = \sum_{i=1}^8 p_{ij} m_{ij}$$

目标函数应为年购买成本最小，即

$$\min C = \sum_{j=1}^4 C_j = \sum_{j=1}^4 \sum_{i=1}^8 p_{ij} m_{ij} \quad (8.1)$$

8.1.2 约束条件的确定

(1) 营养素的保证

用 x_{ik} 表示每 100g 第 i 种水果中第 k 种营养素的含量，从零售市场到餐桌的损

耗率为 α_{i3} ，假设每个季度有 91 天，则针对第 j 季度，第 k 种营养素，有

$$91S_{k \min} \leq 10(1 - \alpha_{i3})x_{ik}m_{ij} \leq 91S_{k \max} \quad (8.2)$$

上式中的系数 10 是为了将蔬菜购买的质量(kg)转化为附件中提供的单位(100g)。

(2) 变量取值约束

所有变量均为正数，即

$$p_{ij} \geq 0; m_{ij} \geq 0 \quad (8.3)$$

综上所述，得到基于成本最小的主要果蔬购买策略的优化

$$\begin{aligned} \min \quad & C = \sum_{j=1}^4 C_j = \sum_{j=1}^4 \sum_{i=1}^8 p_{ij} m_{ij} \\ s.t. \quad & \begin{cases} 91S_{k \min} \leq 10(1 - \alpha_{i3})x_{ik}m_{ij} \leq 91S_{k \max} \\ p_{ij} \geq 0; m_{ij} \geq 0 \end{cases} \end{aligned} \quad (8.4)$$

8.2 模型的求解

8.2.1 不同区域价格 p_{ij} 的确定

由于价格 p_{ij} 数据非常不齐全，通过查找大量的网站，最终确定以数据相对较全的金投网^[11]和蔬菜商情网^[12]为数据蓝本，选取河北省石家庄市，江苏省吴州市，和山东省威海市的数据作为代表进行模型的求解，且当该地数据缺失时，以其相邻市的数据代替。

根据文献[11]提供的数据，整理得到自 2013 年 12 月份至 2014 年 9 月份的月均价格^③，如下表所示。

表 16：三个地区主要水果价格(单位：元/kg)^④

	苹果			梨			香蕉			葡萄		
	石家庄	吴忠	威海	石家庄	吴忠	威海	石家庄	吴忠	威海	石家庄	吴忠	威海
2013.10	--	--	--	--	--	--	--	--	--	--	--	--
2013.11	--	--	--	--	--	--	--	--	--	--	--	--
2013.12	3.3	4.59	4.4	4.45	4.23	4	4.5	5.22	4	9	12	8
2014.01	3.3	-	5	4.45	4.23	4	4.5	5	3	9	--	--
2014.02	--	--	--	4.45	4.23	4	--	--	--	--	--	16
2014.03	4.9	5	4.4	5	7.3	5	--	6	--	--	8.6	12
2014.04	5.5	5	7	5.6	8.76	5	2.8	6.5	7.8	--	--	20
2014.05	6.5	5	5	6.6	9.96	6	2	8.22	5.6	20	13	--
2014.06	6.1	5	7	8	8.76	7.4	4	--	6.5	--	--	--
2014.07	2.7	5	5	5	7.76	7.2	--	4.86	6.5	--	8	11
2014.08	3	5	5	5	9.36	7.2	4	3.3	3.6	4.4	5.6	9
2014.09	3.3	4.6	4	9.2	6.75	8	2	5.36	--	4.4	5	8.5

^③ 该价格均为批发价格，具体零售价格未能查到。

^④ “--”表示未查到数据，也未查到邻近省市的替代数据，故空缺，下同

表 17：三个地区主要蔬菜价格(单位：元/kg)^⑤

	花菜			黄瓜			香菇			胡萝卜		
	石家庄	吴忠	威海	石家庄	吴忠	威海	石家庄	吴忠	威海	石家庄	吴忠	威海
2013.10	--	--	--	--	--	--	--	--	--	--	--	--
2013.11	--	--	--	--	--	--	5.3	5.3	5.3	--	--	--
2013.12	4.59	6	4.59	3.4	4.4	2.7	8	8	8	2	3	2
2014.01	--	--	4.59	3.4	4.4	2.7	8.4	8.4	8.4	2.8	--	2.2
2014.02	5.9	--	4.59	3.4	4.4	2.7	--	--	--	2.8	--	2.2
2014.03	3.2	6.6	4	3.6	3	4	10	10	10	3.6	2.6	1.8
2014.04	5	5.2	5	1.8	1.8	3.5	45	45	45	2	2.5	2
2014.05	4	7	3	1.4	1.2	1.2	--	--	--	3	3.2	1.8
2014.06	--	5	4	1.2	1.1	1	--	--	--	1.4	2.4	1.4
2014.07	6	5	3.4	1.2	1.6	1.4	5.4	5.4	5.4	1.3	3.4	1.4
2014.08	4	7	5	2.2	2.7	2.4	4.5	4.5	4.5	1	1.2	1.4
2014.09	3.8	6	3.4	1.6	1.6	2	--	--	--	1.2	1	1.4

从上表可以看出，同一地区不同月份价格有所差别，季节性因素较为明显，同一月份不同地区的价格差异性较大，地区性因素也比较明显。因此分地区，忽略未查到的数据取季度价格的平均值作为该季度该水果的平均价格，则有第 i 种果蔬第 j 个季度的平均价格 p_{ij} 取值如下表所示。

表 18：第 i 种果蔬第 j 个季度的平均价格 p_{ij} 取值表(单位：元/kg)

	第一季度			第二季度			第三季度			第四季度		
	石家庄	吴忠	威海	石家庄	吴忠	威海	石家庄	吴忠	威海	石家庄	吴忠	威海
苹果	4.1	5	4.7	6.03	5.00	6.33	3.15	4.87	4.67	3.3	4.59	4.4
梨	4.63	5.25	4.33	6.73	9.16	6.13	6.40	7.96	7.47	4.45	4.23	4
香蕉	4.5	5.5	3	2.93	7.36	6.63	3.00	4.51	5.05	4.5	5.22	4
葡萄	9	8.6	14	20	13	20	4.40	6.20	9.50	9	12	8
花菜	4.55	6.6	4.39	4.5	5.73	4.00	4.60	6.00	3.93	4.59	6	4.59
黄瓜	3.47	3.93	3.13	1.47	1.37	1.90	1.67	1.97	1.93	3.4	4.4	2.7
香菇	9.2	9.2	9.2	45.00	45.00	45.00	4.95	4.95	4.95	6.65	6.65	6.65
胡萝卜	3.07	2.6	2.07	2.13	2.70	1.73	1.17	1.87	1.40	2	3	2

8.2.2 模型结果的求解

根据第 1,2 问数据，结合表(16)及模型(8.4)，利用 LINGO 编程，得到三个地区居民的最优购买策略如下表所示。

^⑤ 香菇的价格非常稀少，故以平均价格作为三个市的对应价格，因此三个市的香菇价格在不同月份是一样的。

表 19：石家庄居民最优购买策略(单位：kg 和元)

季度	苹果	梨	香蕉	葡萄	花菜	黄瓜	香菇	胡萝卜	总花费
一	127.5	69.0	48.0	72.0	42.6	63.2	50.4	55.5	2753.4
二	166.5	90.0	61.5	94.5	55.6	81.9	67.4	72.0	7236.0
三	139.5	85.5	60.0	73.5	52.8	78.6	46.2	54.0	2156.1
四	138.0	78.0	56.5	80.0	48.2	72.7	53.3	58.0	2715.6
年度	571.5	322.5	226.0	320.0	199.3	296.5	217.2	239.5	14861.1

表 20：吴忠居民最优购买策略(单位：kg 和元)

季度	苹果	梨	香蕉	葡萄	花菜	黄瓜	香菇	胡萝卜	总花费
一	150.0	75.0	42.0	88.5	46.4	63.2	75.2	61.5	3541.6
二	175.5	97.5	52.5	114.0	60.3	81.1	98.7	61.5	8704.9
三	165.0	78.0	49.5	106.5	48.2	68.9	86.8	58.5	3272.0
四	154.5	79.5	48.0	103.0	57.0	68.9	77.5	51.5	3847.1
年度	645.0	330.0	192.0	412.0	211.8	282.2	338.2	233.0	19365.6

表 21：威海居民最优购买策略(单位：kg 和元)

季度	苹果	梨	香蕉	葡萄	花菜	黄瓜	香菇	胡萝卜	总花费
一	139.5	87.0	48.0	82.5	53.8	73.0	67.7	52.5	3527.6
二	180.0	114.0	64.5	108.0	70.5	96.5	87.0	66.0	8922.3
三	148.5	106.5	67.5	88.5	65.8	94.1	61.7	42.0	3475.0
四	145.5	105.0	60.0	93.0	75.0	89.2	63.0	40.5	3129.2
年度	613.5	412.5	240.0	372.0	265.0	352.7	279.5	201.0	19054.2

8.3 模型的改进：基于居民消费价格指数 CPI 模型的修正

模型(8.4)考虑的是一种静态最优购买策略，对于同一种果蔬而言，即使考虑同一个季度，10 年前的单价和 10 年后还是有一定差异的，因为随着我国经济的发展，物价也在不断上涨。居民消费指数 CPI 反映了居民家庭购买消费商品及服务的价格水平的变动情况，因此可以根据 CPI 对模型(8.4)进行动态修正。

以 2013 年为例，考虑第 i 年($i > 2013$)的最优购买策略，设 i 年相对于 2013 年第 j 个季度的 CPI 为 CPI_{ij} ，则模型(8.4)的约束条件不变，修正后的模型为

$$\begin{aligned}
 \min \quad & C \sum_{j=1}^4 \sum_{i=1}^8 CPI_{ij} p_{ij} m_{ij} \\
 s.t. \quad & \begin{cases} 91S_{k \min} \leq 10x_{ik} m_{ij} \leq 91S_{k \max} \\ p_{ij} \geq 0; m_{ij} \geq 0 \end{cases}
 \end{aligned} \tag{8.5}$$

结合相关数据即可确定动态意义下模型(8.5)的最优解，限于篇幅与实践，本文不再预测 2014-2020 年的季度 CPI 并进行模型的求解。

8.4 结果分析

表 19-表 21 表明如下结论：

- 为了摄入的营养素达到保障人体健康所必须的范围，不同地区的总花费是不同的，这和当地的消费水平有关，一般而言，消费水平越高，花费越多，反之亦然；
- 结果看上去似乎有点奇怪，以石家庄市苹果为例，人均年苹果消耗量达到 571.5kg，平均日进食 1.57kg 的苹果，看上去似乎有点大，由于本文仅考虑这 8 类主要果蔬，未考虑肉类食物、主要粮食的进食，因此结果看上去偏大是正常的。
- 第二季度的花费要高于上下 3 个季度的花费，这主要是由于第二个季度的香菇价格远高于其他季度，以及第二季度其余主要果蔬的平均价格也偏高的原因。

综上所述，本模型能较好地给与居民提供合适价廉的购买建议。

9、问题 4 模型的建立与求解

9.1 问题分析与符号说明

本问将在尽量满足营养摄入量的前提下，综合考虑居民的购买成本、种植者的利润以及进出口贸易和土地面积等制约因素，建立多目标优化模型确定我国居民主要果蔬产品的年度人均消费量，本问涉及到众多变量，因此单独给出一张符号说明表。

符号(单位)	含义
p_i (元/kg)	第 i 种果蔬的年均零售价， $i=1,2,3,4,5$ 分别表示苹果，梨，葡萄，香蕉和蔬菜 ^⑥ ，下同
c_i (元/kg)	第 i 种果蔬的年均成本
$\gamma_i = \tau_i S_i$ (kg)	第 i 种果蔬年均种植产量
m_i (kg)	第 i 种果蔬居民年均购买量
x_{ik}	第 i 种果蔬每 100g 中第 k 种营养素的含量
EX_i (kg)	第 i 种果蔬年出口量
IM_i (kg)	第 i 种果蔬年进口量
α_{i1}, α_{i3}	第 i 种果蔬从田间地头到大市场的损耗率，从零售市场到餐桌的损耗率

^⑥ 根据中华人民共和国国家统计局数据，蔬菜作为主要农作物中的一类，未对其进行细分，因此无法查找蔬菜类别下菜花，黄瓜，香菇，胡萝卜的数据，因此本问将蔬菜作为一大类予以考虑。

9.2 多目标优化模型的建立

9.2.1 目标函数的确定

(1) 居民的购买成本

居民年均购买第 i 种果蔬的质量为 m_i (kg)，则实际购买成本(单位：元)

$$F_1 = \sum_{i=1}^5 p_i m_i \quad (9.1)$$

(2) 种植者的利润

种植者的年产量为 γ_i ，从田间地头到大市场的的损耗率为 α_{i1} ，因此种植者实际可卖产量为 $(1-\alpha_{i1})\gamma_i$ ，种植者的实际利润(单位：元)

$$F_2 = \sum_{i=1}^5 (p_i - c_i)(1-\alpha_{i1})\gamma_i \quad (9.2)$$

(3) 出口贸易量

出口贸易是对外贸易的基础，作为一个农业大国，在保证我国居民营养均衡的条件下，希望出口贸易量 EX_i 越大越好。

9.2.2 约束条件的确定

(1) 营养素保证的约束

居民购买第 i 种果蔬质量为 m_i kg，从零售市场到餐桌的损耗率为 α_{i3} ，所以实际摄入的果蔬质量为 $(1-\alpha_{i3})m_i$ 。用 x_{ik} 表示每 100g 第 i 种水果中第 k 种营养素的含量，考虑一年 365 天，则为了保证年均营养素的摄入，有

$$365S_{k\min} \leq 10 \sum_{i=1}^5 (1-\alpha_{i3})x_{ik}m_i \leq 365S_{k\max} \quad \forall k = 1, 2, \dots, 20 \quad (9.3)$$

不等式中间部分系数 10 是为了保证不等式左中右部分的质量单位均为 100g。

(2) 土地面积约束(进出口约束)

土地面积约束即为居民购买质量不超过国家总的供应量，而总供应量又等于减去损耗之后的总产量-出口量+进口量，即

$$\sum_{i=1}^5 m_i \leq (1-\alpha_{i1})\gamma_i - \sum_{i=1}^5 EX_i + \sum_{i=1}^5 IM_i \quad (9.4)$$

9.2.3 优化模型的确定

综合 9.2.1-9.2.2，得到多目标优化模型

$$\begin{aligned}
\min F_1 &= \sum_{i=1}^5 p_i m_i \\
\max F_2 &= \sum_{i=1}^5 (p_i - c_i)(1 - \alpha_{i1}) \gamma_i \\
\max F_3 &= EX_i \\
s.t. &\begin{cases} 365S_{k\min} \leq 10 \sum_{i=1}^5 (1 - \alpha_{i3}) x_{ik} m_i \leq 365S_{k\max} \\ \sum_{i=1}^5 m_i \leq (1 - \alpha_{i1}) \gamma_i - \sum_{i=1}^5 EX_i + \sum_{i=1}^5 IM_i \\ EX_i, IM_i \geq 0 \\ k = 1, 2, \dots, 20; i = 1, 2, 3, 4, 5 \end{cases}
\end{aligned} \tag{9.5}$$

9.3 多目标优化模型的求解

为了书写和叙述上的方便，上述建模过程并未涉及到具体年份，实际上，不同的年份上述参数和变量取值均不同，在模型的求解中，首先确定(查找或预测)模型(9.5)的参数在 2011-2020 年的取值，进而求解模型，确定到 2020 年我国果蔬产品生产的调整战略。

9.3.1 模型的参数确定

(1) 2011-2020 年年均产量 γ_i 的确定

中国统计局给出了 1995-2011 年(2012-2014 年数据不全)的年均产量 γ_i (数据见附录 E)，利用 MATLAB 做出年均产量的变化示意图，如图 10 和图 11 所示。

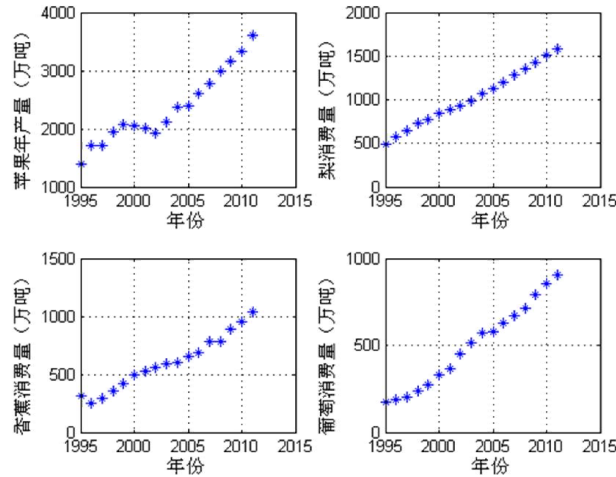


图 10: 水果年均产量变化图

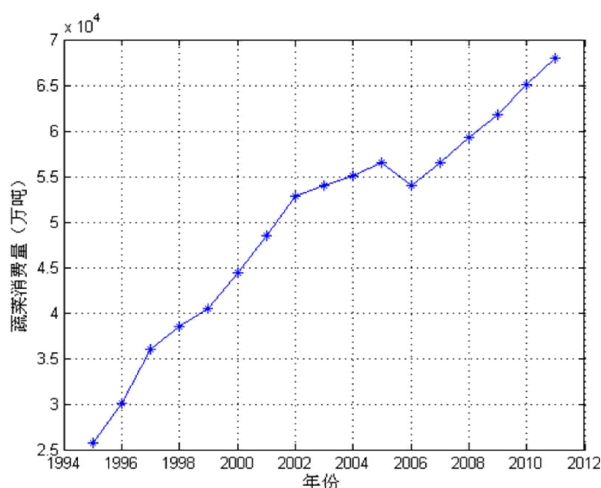


图 11: 蔬菜年均产量变化图

由图 10 和图 11 可知, 自 2006 年以后, 水果和蔬菜的增长呈“线性增长”趋势, 因此根据 2006-2011 年数据建立线性回归模型预测 2012-2020 年年均产量。

用 γ_{ij} 表示第 i 种果蔬第 j 年的年均产量, 其中 $i=1, \dots, 5; j=2006, \dots, 2020$, 建立线性回归模型

$$\gamma_{ij} = b_{i1}i + b_{i2} + \varepsilon$$

其中 b_{i1} 和 b_{i2} 为系数, $\varepsilon \sim N(0,1)$ 为随机误差。

利用 MATLAB, 根据最小二乘法确定 5 种果蔬的系数及检验结果如下表所示。

表 22: 线性回归模型结果及其检验

	苹果	梨	香蕉	葡萄	蔬菜
b_1	193.4	74.97	67.96	58.08	2812
b_2	-3.853×10^5	-1.492×10^5	-1.356×10^5	-1.159×10^5	-5.587×10^6
检验结果 R^2	0.9937	0.9981	0.9639	0.9897	0.9980

检验结果 R^2 非常接近 1, 说明拟合效果很小, 拟合结果可用。根据上述结果, 得到 2012-2020 年果蔬年均产量如所示

表 23:2012-2020 年果蔬年均产量预测值(万吨)

年份	苹果	梨	香蕉	葡萄	蔬菜
2012	3820.8	1639.6	1135.5	956.96	70744
2013	4014.2	1714.6	1203.5	1015	73556
2014	4207.6	1789.6	1271.4	1073.1	76368
2015	4401	1864.5	1339.4	1131.2	79180
2016	4594.4	1939.5	1407.4	1189.3	81992
2017	4787.8	2014.5	1475.3	1247.4	84804
2018	4981.2	2089.5	1543.3	1305.4	87616
2019	5174.6	2164.4	1611.2	1363.5	90428
2020	5368	2239.4	1679.2	1421.6	93240

(2) 2011-2020 年年均零售价 p_i 和年均成本 c_i 的确定

成本与零售价的变化涉及到多种复杂因素，如地域，经济水平，消费观念等等，而数据又非常难稀少，假设年均零售价与年均成本在短期(2020 年之内)没有大的变化，设为定值。

根据表 18，以 3 市各季度果蔬价格的平均值作为对应果蔬年均零售价的估计，结合网络数据^⑦[13-19]，得到 5 种果蔬的年均零售价 p_i 和年均成本 c_i 如下表所示。

表 24:2012-2020 年果蔬年均零售价与年均成本(单位：元/kg)

	苹果	梨	香蕉	葡萄	蔬菜
年均零售价	4.68	5.90	4.68	11.14	6.54
年均成本	1.7	1.6	1.33	3.33	3.01

(3) 2011-2020 年进口量 IM_i 的确定

FAOSTAT 和 USDA 提供了 1995-2011 年果蔬的进口量，图 12 给出了 1995-2011 年果蔬的进口量。

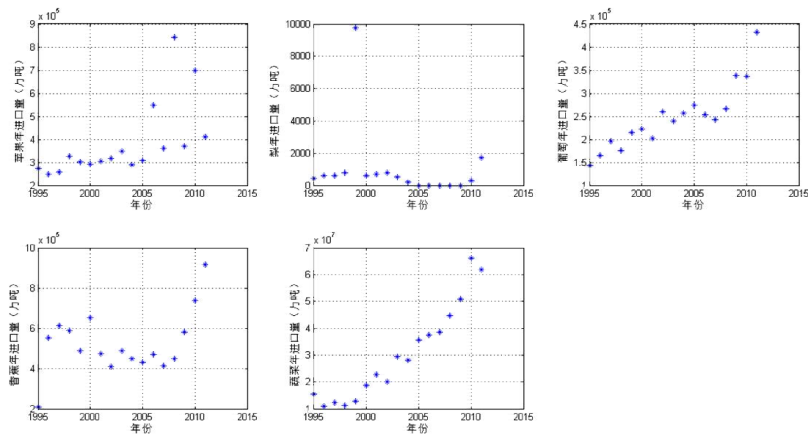


图 12: 1995-2011 年果蔬的进口量

从图 12 可以看出，进口量没有明显的变化规律，走势较为杂乱。因此取 1995-2011 年的平均值作为 2012-2020 年的进口预测值，见表 25。

表 25: 2012-2020 年果蔬的年均出口量(单位：万吨)

	苹果	梨	香蕉	葡萄	蔬菜
年均进口量	382276	1000	248443	524595	30408043

9.3.2 模型的求解

模型(9.5)是三目标优化问题。一般而言，多目标优化问题没有全局意义上的最优解，目标加权法和分层排序法是常用的多目标优化模型的求解方法。基于“以

^⑦ 对于同一果蔬若出现不同成本，取平均值。

民为本”的思想，本文采用分层排序法对模型(9.5)进行求解。

目标函数 1 为极小型目标函数，目标函数 2 为极大型目标函数且两个目标的量纲相同，将两个目标函数进行组合，有

$$\max F = F_2 - F_1 = \sum_{i=1}^5 ((p_i - c_i)(1 - \alpha_{i1})\gamma_i - p_i m_i) \quad (9.6)$$

结合(9.5)和(9.6)，得到新的双目标优化模型

$$\begin{aligned} \max F &= \sum_{i=1}^5 ((p_i - c_i)(1 - \alpha_{i1})\gamma_i - p_i m_i) \\ \max F_3 &= EX_i \\ s.t. &\begin{cases} 365S_{k\min} \leq 10 \sum_{i=1}^5 (1 - \alpha_{i3}) x_{ik} m_i \leq 365S_{k\max} \\ \sum_{i=1}^5 m_i \leq (1 - \alpha_{i1})\gamma_i - \sum_{i=1}^5 EX_i + \sum_{i=1}^5 IM_i \\ EX_i, IM_i \geq 0 \\ k = 1, 2, \dots, 20; i = 1, 2, 3, 4, 5 \end{cases} \end{aligned} \quad (9.7)$$

基于“以民为本”的思想，在模型求解中，首先求目标函数 F 的最优解，在目标函数 F 的最优解的参数取值下，再求 F_3 的最优解。

利用 MATLAB 进行编程，得到 2012-2020 年每年第 i 种果蔬的购买量如下：

表 26: 国家宏观战略下 2012-2020 年果蔬的人均理性消费量 (单位: kg)

	苹果	梨	葡萄	香蕉	蔬菜
2012	650	365	267	304	1285
2013	657	361	259	301	1297
2014	689	382	279	307	1294
2015	688	383	276	305	1304
2016	702	389	278	315	1307
2017	700	390	275	318	1302
2018	685	382	267	312	1330
2019	687	374	273	317	1298
2020	673	383	277	318	1305
平均值	681	379	272	311	1302

9.4 结果分析

根据表 26 的结论可知：

- 到 2020 年，我国人均果蔬平均消耗量已基本稳定，5 类主要果蔬，消费量从高到低依次是：蔬菜、苹果，梨，香蕉，葡萄；
- 由于膳食纤维只能通过蔬菜来提供，因此蔬菜的人均消费量最多；
- 由于本文没有考虑粮食和肉类食物的摄入量，因此需要所计算的结果与直觉上相比较偏大，这是合理的。

10、基于模型结果的政策建议：给相关部门的建议

尊敬的领导：

我们是来自**学校的研究生，对人体营养健康角度的中国果蔬发展战略进行了研究，先根据我们的研究结论，结合实际情况，从种植品种选择、引导居民理性消费、进出口贸易选择三个方面向您提供一些政策建议。

一、合理地选择种植品种

根据问题 1 中我们基于营养素成分和含量的聚类模型显示，不同年龄段的人群应适当偏向于使用不同的果蔬，具体情况如下表所示。

表 27：不同年龄段人群推荐果蔬品种

年龄段	推荐水果品种
0-4 岁	樱桃，柠檬，橘子，木瓜，草莓，番荔枝，椰子，大枣
5-18 岁	苹果，梨，草莓，葡萄，火龙果，荔枝，哈密瓜，柠檬
19-45 岁	苹果，梨，葡萄，火龙果，荔枝，无花果，琵琶，猕猴桃
46 岁及以上	苹果，无花果，琵琶，猕猴桃，香蕉，哈密瓜，桑葚，金橘
年龄段	推荐蔬菜品种
0-4 岁	苦瓜，南瓜，黄瓜，圆白菜，青椒，菠菜，菜花，蘑菇
5-18 岁	生菜，西红柿，黄瓜，圆白菜，菠菜，菜花，胡萝卜，蘑菇
19-45 岁	竹笋，芹菜，大白菜，蒜苗，丝瓜，竹笋，韭菜，菜花，香菇
46 岁及以上	竹笋，韭菜，土豆，蒜苗，丝瓜，大白菜，韭菜，菜花，香菇

因此，我们建议您根据不同地区的年龄结构组成在当地种植或销售对应的果蔬，一来可以较好地为居民提供恰当的营养，二来也可以获得较大的经济收入。

二、引导居民理性消费

根据我们问题 2 的结论，我国居民目前矿物质的年摄入水平已达到保障人体健康所需营养素成分的参考范围；维生素的年摄入量略微偏离参考范围；而膳食纤维的年摄入水平呈缓慢下降趋势。

整体上而言，我国居民营养素的年摄入水平良好，反映了人们对人体均衡营养的意识有所增强。然而，多数中国居民喜食、饱食、偏食等传统不良饮食习惯并没有改变。

根据我们问题 3 的结论，我们建议您宏观调控不同地区的主要果蔬价格，使得当地居民所应着重补充的营养素对应的果蔬价格相对较低，这样可以合理地引导居民进行理性消费；同时还应该增加科普宣传力度，提高居民的健康营养意识。

三、调控进出口贸易

根据我们问题 4 的结论，为了考虑我国居民的营养均衡，估计居民的购买成本，并且考虑种植者的收益，从宏观调控的角度，进出口贸易需要有一定的额度，尤其是出口贸易。

由于我国人口众多，是世界上的农业大国，我们建议在进行宏观调控的时候，优先满足我国居民的营养摄入量达标；控制主要果蔬的价格，照顾居民的购买成本；加大农作物种植补贴，使得种植者能够获得较大的收益。

此致

敬礼

研究生： **

2014 年 9 月 23 日

11、优缺点分析

11.1 优点分析

- 问题 1 中针对不同年龄段的人群筛选了不同水果和蔬菜，结果更具有针对性；
- 问题 1 中先判断数据走向再选用 GM(1,1)模型，使得模型的运用更具有说服力；
- 问题 2 中将健康水平分层化，使得结果更容易被理解和接受；
- 问题 3 中分地区分季度考虑价格变化，提高了模型的准确性；
- 问题 3 根据 CPI 指数对优化模型进行了改进，考虑到了物价上涨对结果带来的影响；
- 问题 4 采用分层排序法及目标整合法对多目标优化模型进行了化简，降低了模型求解的难度。

11.2 缺点分析

- 由于数据量的匮乏，在问题 1 中筛选主要果蔬品种时，筛选结果较小，在一定程度上降低了模型结果的精确度；
- 问题 1 中 GM(1,1)模型只能做短期预测，长期预测效果不好(在问题 2 的结果中有所体现)；
- 问题 3 由于数据不全，仅能给出部分地区的购买建议；
- 问题 4 由于数据不全，参数取值精确性有待进一步提高。

参考文献

- [1] 王学民. 应用多元统计 (第三版) [M]. 上海: 上海财经大学出版社, 2011.
- [2] 司守奎, 孙玺青. 数学建模算法与应用[M]. 北京: 国防工业出版社, 2011.
- [3] 杨月欣, 王光亚, 潘兴昌. 中国食物成分表[M]. 北京: 北京大学医学出版社, 2002.
- [4] 韩中庚. 数学建模及其应用(第二版)[M]. 北京: 高等教育出版社, 2012.
- [5] 姜启源, 谢金星. 数学模型(第三版)[M]. 北京: 高等教育出版社, 2008.
- [6] 洛建忠. 基于营养目标的粮食消费需求研究[D]. 北京: 中国农业科学院, 2008.
- [7] 张峭, 杨霞. 中国水果消费现状分析及其预测[J]. 农业展望, 2006, 8: 30-33.
- [8] 李哲敏. 中国城乡居民食物消费与营养发展的趋势预测分析[J]. 农业技术经济, 2008, 6:57-62.
- [9] 中国经济网. 数据显示我国居民膳食纤维摄入量呈下降趋势.
<http://finance.sina.com.cn/roll/20140523/154219208430.shtml?qq-pf-to=pcqq.c2>, 2014-9-20.
- [10] 国家统计局. 总人口. <http://data.stats.gov.cn/workspace/index?m=hgnd>, 2014-9-21.
- [11] 金投网. 水果/蔬菜价格. <http://jiage.cngold.org/shuiguao/>, 2014-9-21.
- [12] 蔬菜商情网. 香菇价格. <http://www.shucai123.com/price/p/香菇/>, 2014-9-21.
- [13] 烟台晚报. 苹果种植成本要 1.5 元一斤.
http://news.gmw.cn/newspaper/2013-06/21/content_1622166.htm, 2014-9-22
- [14] 江门日报. 种植成本不到 0.4 元/公斤苹果 卖 12 元/公斤.
http://www.jmnews.com.cn/c/2010/11/19/09/c_1105166.shtml, 2014-9-22

- [15] 中国江西网. 增产难增收, 上万公斤梨子烂地头.
http://epaper.jxnews.com.cn/jxrb/html/2014-07/14/content_252445.htm, 2014-9-22
- [16] 葡萄种植成本和产量. <http://wenda.so.com/q/1385527998065206?src=150>, 2014-9-22.
- [17] 广西农业信息网. 走出香蕉高成本种植的误区.
<http://www.gxny.gov.cn/web/2009-12/266194.htm>, 2014-9-22.
- [18] 算一算大棚蔬菜的种植成本收益账.
http://blog.sina.com.cn/s/blog_834d07400100smez.html, 2014-9-22.
- [19] 大棚蔬菜种植成本每亩的种植需要的成本是多少.
<http://wenda.so.com/q/1371343595061383>, 2014-9-22.

附录

附录 A：表达式(6.1)定义合理性的说明

由于营养素摄入量应尽量在合理范围之内，因此属于居中型指标，对于样本指标 x_{ik} ，越接近第 δ 类人群第 k 项营养素的参考值 $s_k(\delta)$ 越好，但是考虑到针对任一类人群，不同营养素的参考值差异性较大，因此采用无量纲化的相对百分比作为样本与参考值的距离，即

$$x_{ik}^*(\delta) = \frac{|x_{ik} - s_k(\delta)|}{s_k(\delta)}, \quad (\text{A.1})$$

则 x_{ik}^* 是一个极小性指标，为了符合日常习惯（越大越好），因此进一步修正得到

$$x_{ik}^*(\delta) = \frac{1}{\frac{|x_{ik} - s_k(\delta)|}{s_k(\delta)} + M}, \quad (\text{A.2})$$

其中 $M \geq 0$ 为常数是为了避免由于分母为 0 而导致表达式没有意义，然而通过匹配常见水果/蔬菜与营养参考值，对于单位质量(每 100g)下水果/蔬菜，不存在 i 使得 $|x_{ik} - s_k(\delta)| = 0$ ，因此取 $M = 0$ ，整理(A.2)得到

$$x_{ik}^*(\delta) = \frac{s_k(\delta)}{|x_{ik} - s_k(\delta)|}, \quad (\text{A.3})$$

(A.3)即式(6.1)。

附录 B: 4 种年龄段水果和蔬菜营养素聚类图

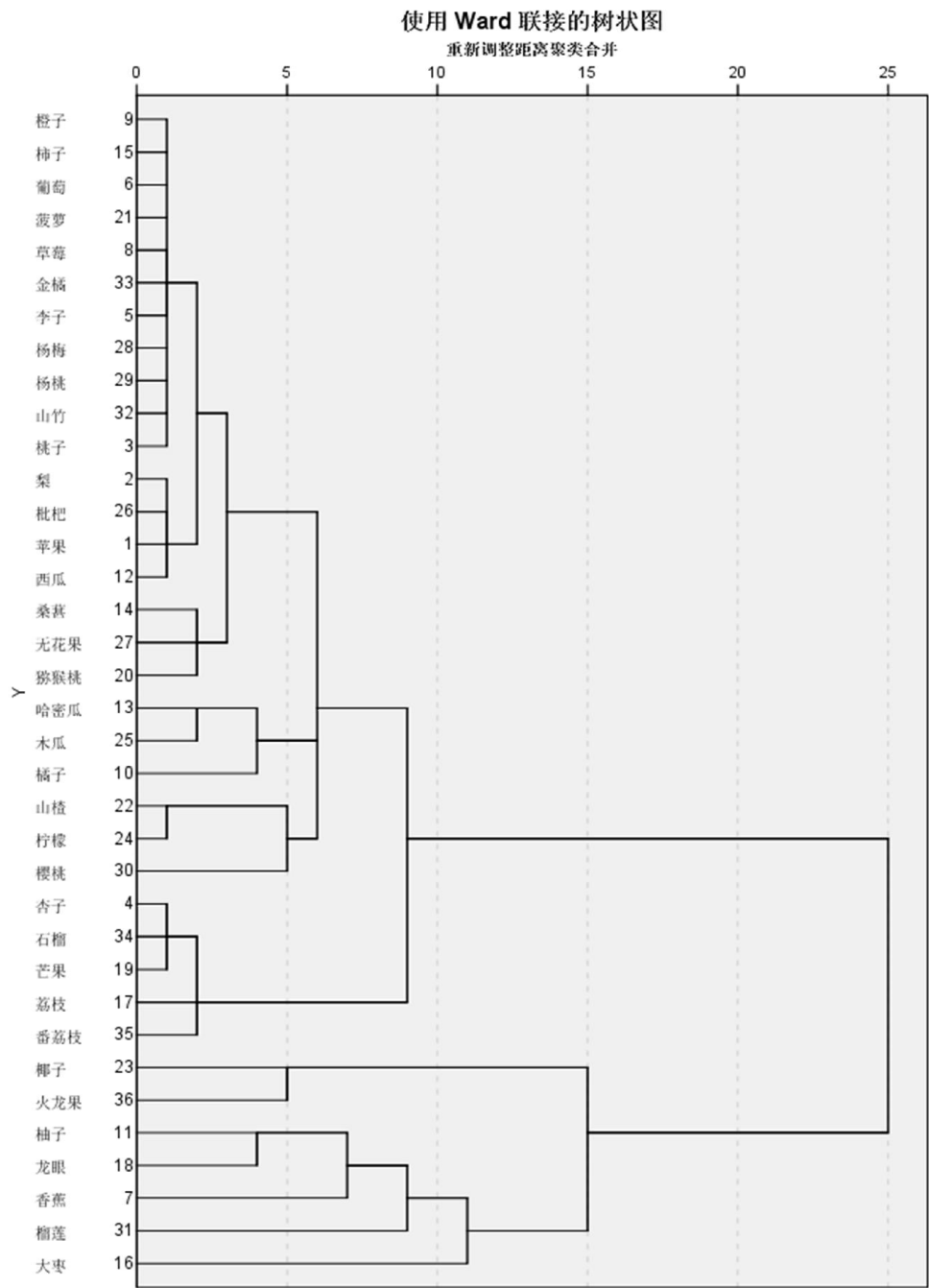


图 13: 0-4 岁水果聚类图

根据聚类树状图，把常见水果营养素对 0--4 岁儿童的效果分为四类：

第一类：樱桃，柠檬，山楂，橘子，木瓜，哈密瓜，猕猴桃，无花果，桑葚，西瓜，苹果，枇杷，梨，桃子，山竹，杨桃，杨梅，李子，金橘，草莓，菠萝，葡萄，柿子，橙子

第二类：番荔枝，荔枝，芒果，石榴，杏子

第三类：火龙果，椰子

第四类：大枣，榴莲，香蕉，龙眼，柚子

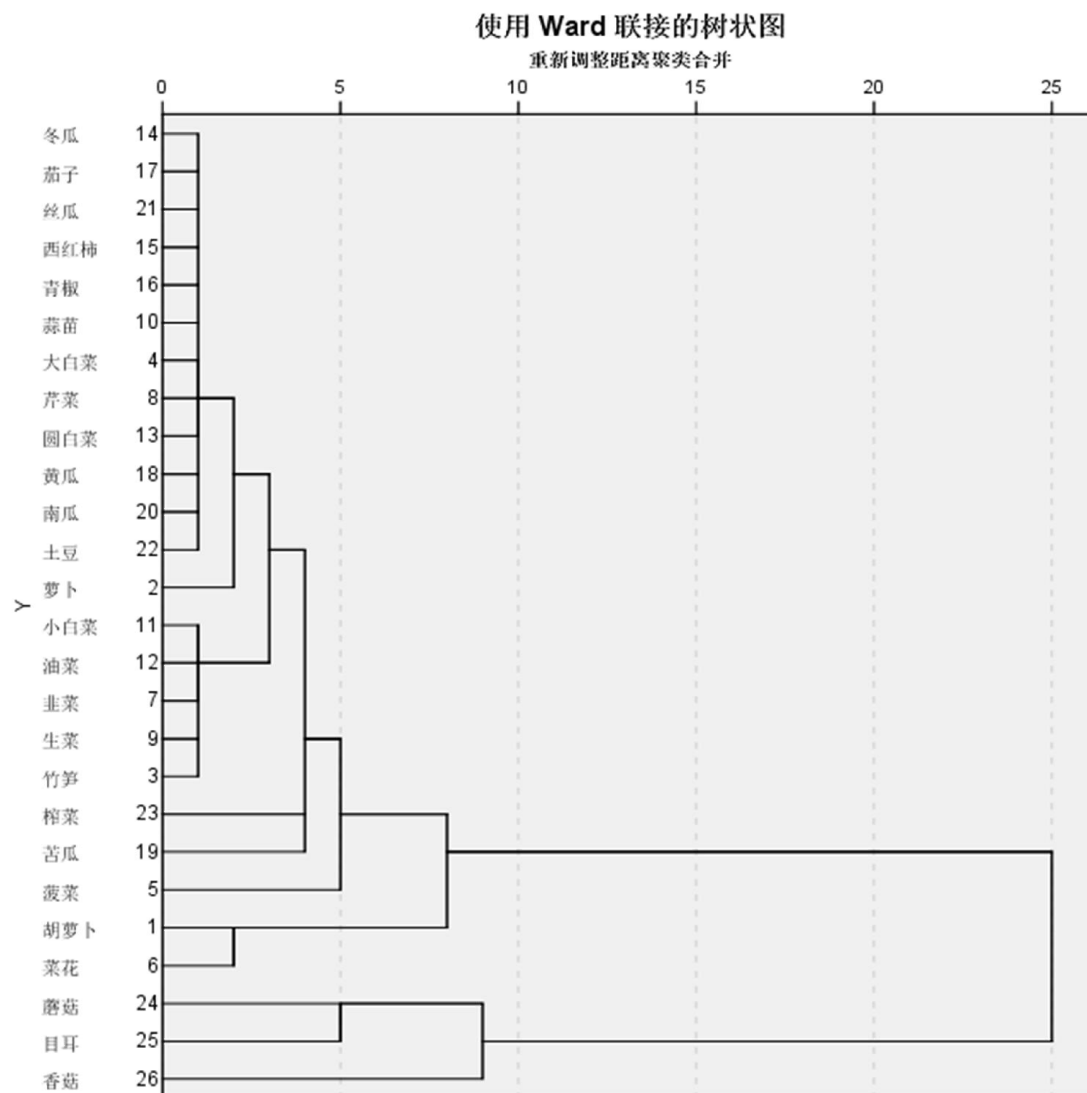


图 14: 0-4 岁蔬菜聚类图

根据聚类树状图，把常见蔬菜营养素对 0-4 岁儿童的效果分为四类：

第一类：苦瓜，榨菜，竹笋，生菜，韭菜，油菜，小白菜，萝卜，土豆，南瓜，黄瓜，圆白菜，芹菜，大白菜，蒜苗，青椒，西红柿，丝瓜，茄子，冬瓜

第二类：菠菜

第三类：菜花，胡萝卜

第四类：香菇，木耳，蘑菇

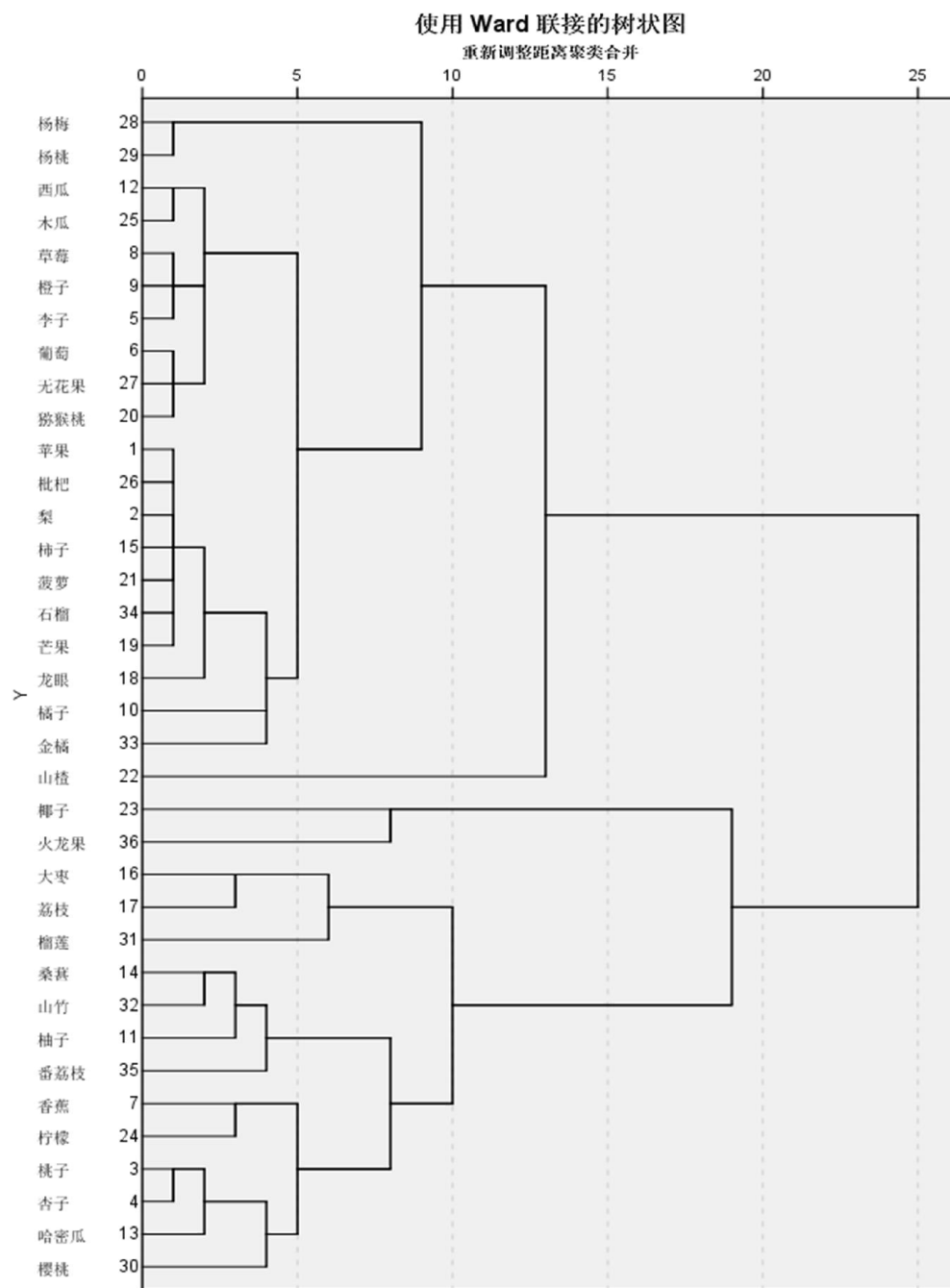


图 15: 5-18 岁水果聚类图

根据聚类树状图，把常见水果营养素对 5--18 岁少年人的效果分为四类：

第一类：山楂，金橘，橘子，龙眼，芒果，石榴，菠萝，柿子，梨，枇杷，苹果，猕猴桃，无花果，葡萄，李子，橙子，草莓，木瓜，西瓜，杨桃，杨梅

第二类：火龙果，椰子

第三类：榴莲，荔枝，大枣

第四类：樱桃，哈密瓜，杏子，桃子，柠檬，香蕉，番荔枝，柚子，山竹，桑葚

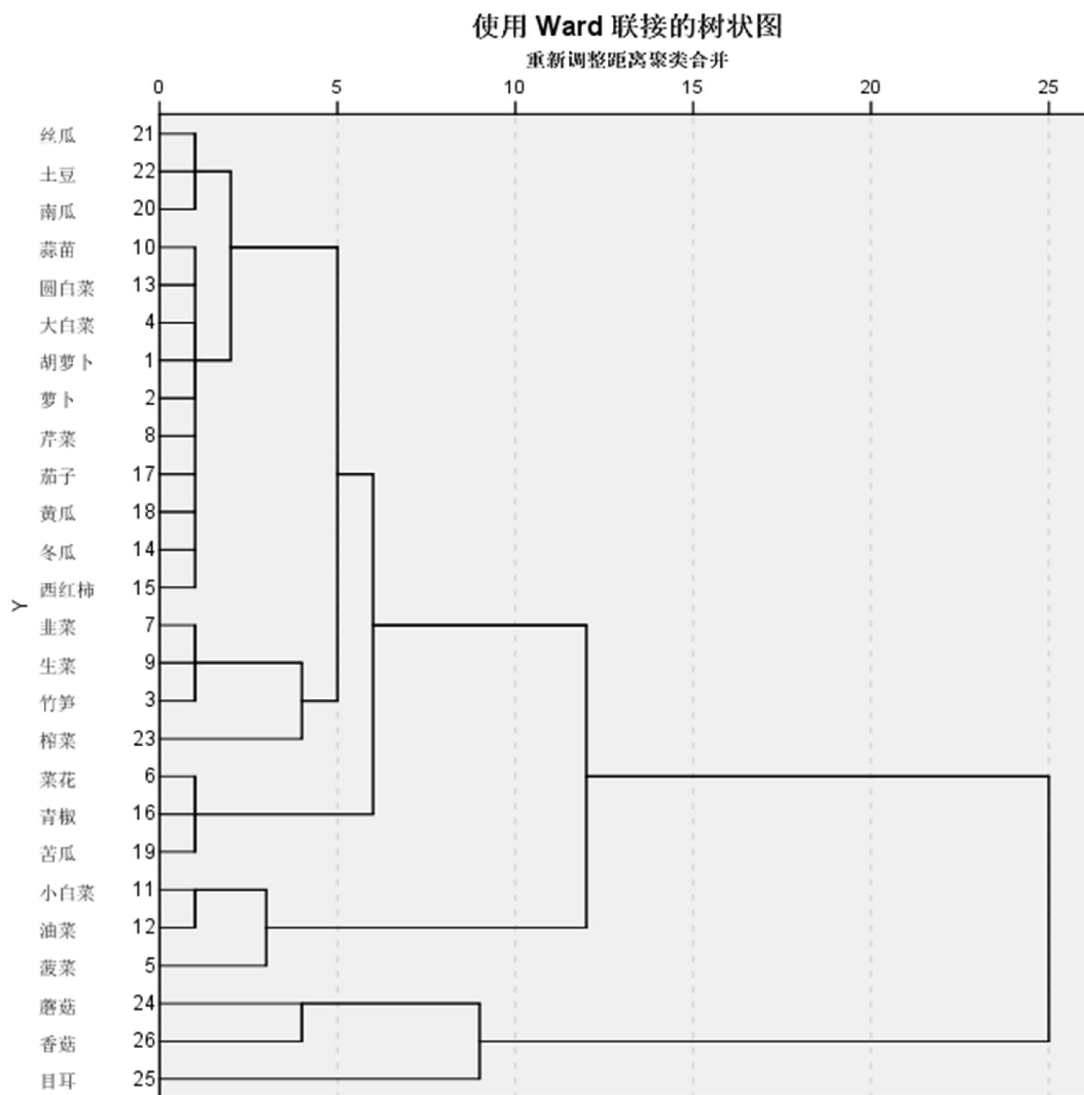


图 16: 5-18 岁蔬菜聚类图

根据聚类树状图，把常见蔬菜营养素对 5--18 岁青少年的效果分为四类：

第一类：榨菜，竹笋，生菜，韭菜，西红柿，冬瓜，黄瓜，茄子，芹菜，萝卜，胡萝卜，大白菜，圆白菜，蒜苗，南瓜，土豆，丝瓜

第二类： 苦瓜，青椒，菜花

第三类：菠菜，油菜，小白菜

第四类： 香菇，木耳，蘑菇

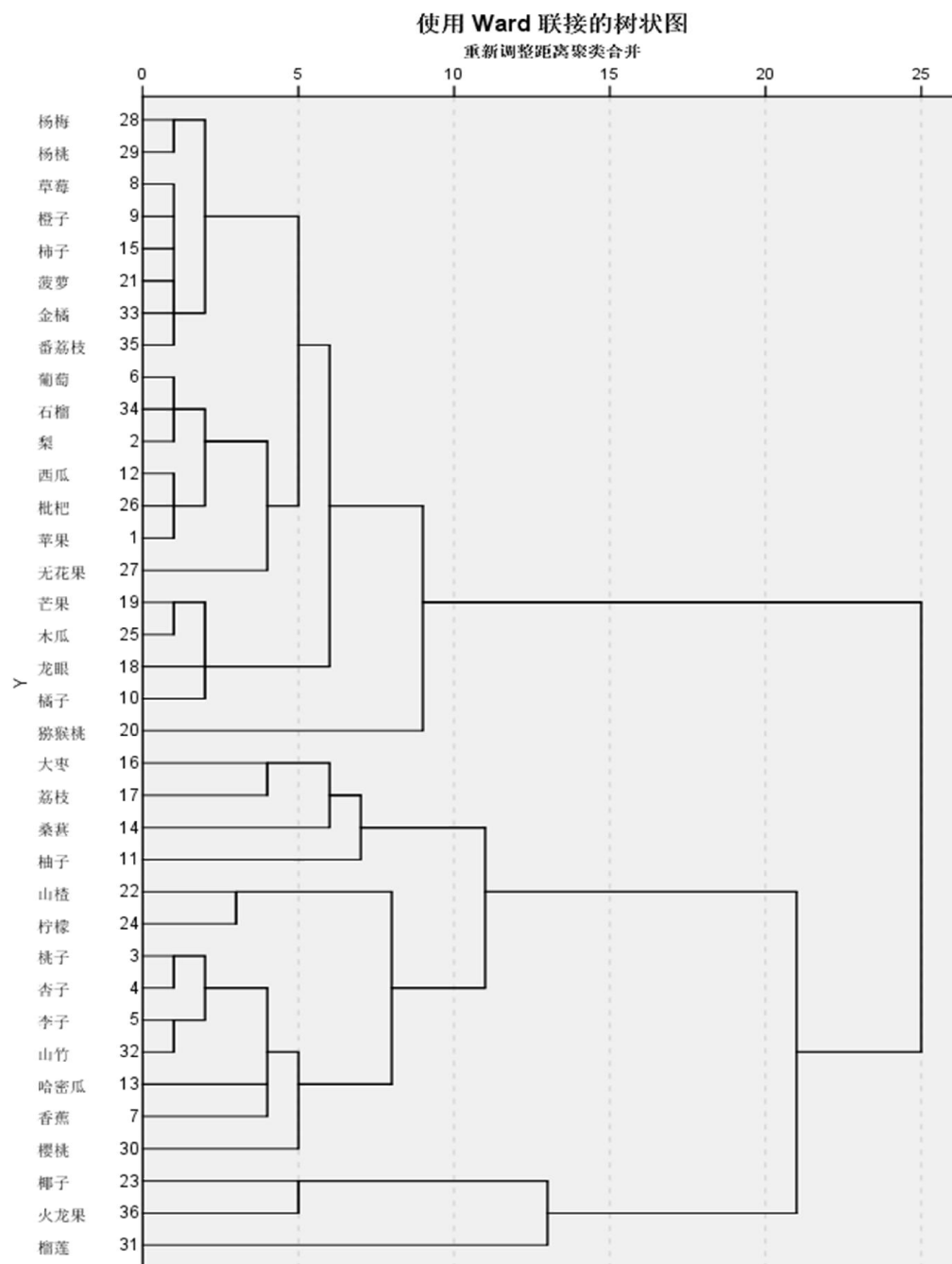


图 17: 19-45 岁水果聚类图

根据聚类树状图，把常见水果营养素对 19--45 岁成年人的效果分为四类：

第一类：橘子，龙眼，木瓜，芒果，无花果，苹果，枇杷，西瓜，梨，石榴，葡萄，荔枝，金橘，菠萝，柿子，橙子，草莓，杨桃，杨梅

第二类：猕猴桃

第三类：樱桃，香蕉，哈密瓜，山竹，李子，杏子，桃子，柠檬，山楂柚子，桑葚，荔枝，大枣

第四类：榴莲，火龙果，椰子

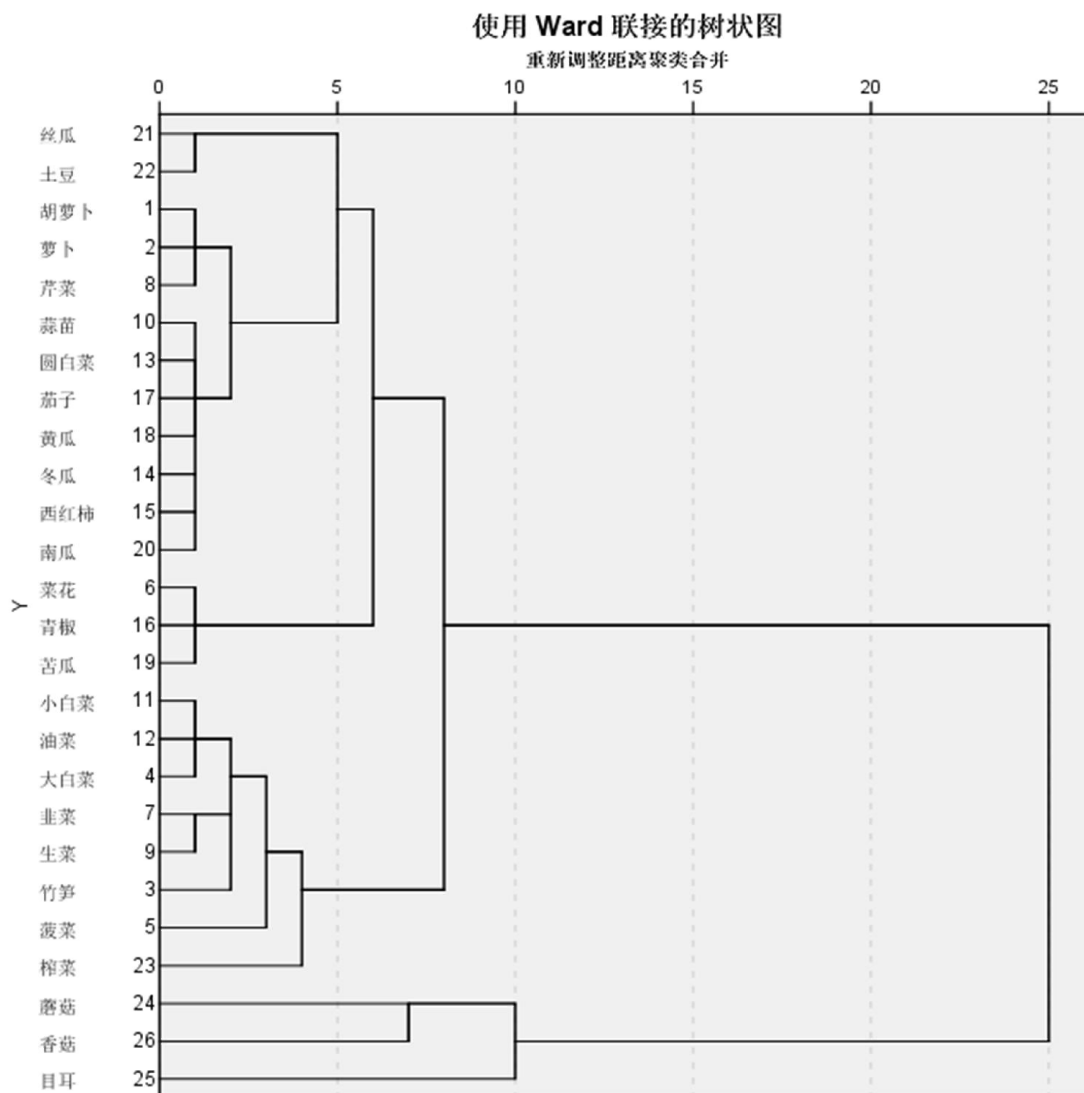


图 18: 19-45 岁蔬菜聚类图

根据聚类树状图，把常见蔬菜营养素对 19--45 岁成年人的效果分为三类：

第一类：苦瓜，青椒，菜花，南瓜，西红柿，冬瓜，黄瓜，茄子，圆白菜，蒜苗，芹菜，萝卜，胡萝卜，土豆，丝瓜

第二类：榨菜，菠菜，竹笋，生菜，韭菜，大白菜，油菜，小白菜

第三类：香菇，木耳，蘑菇

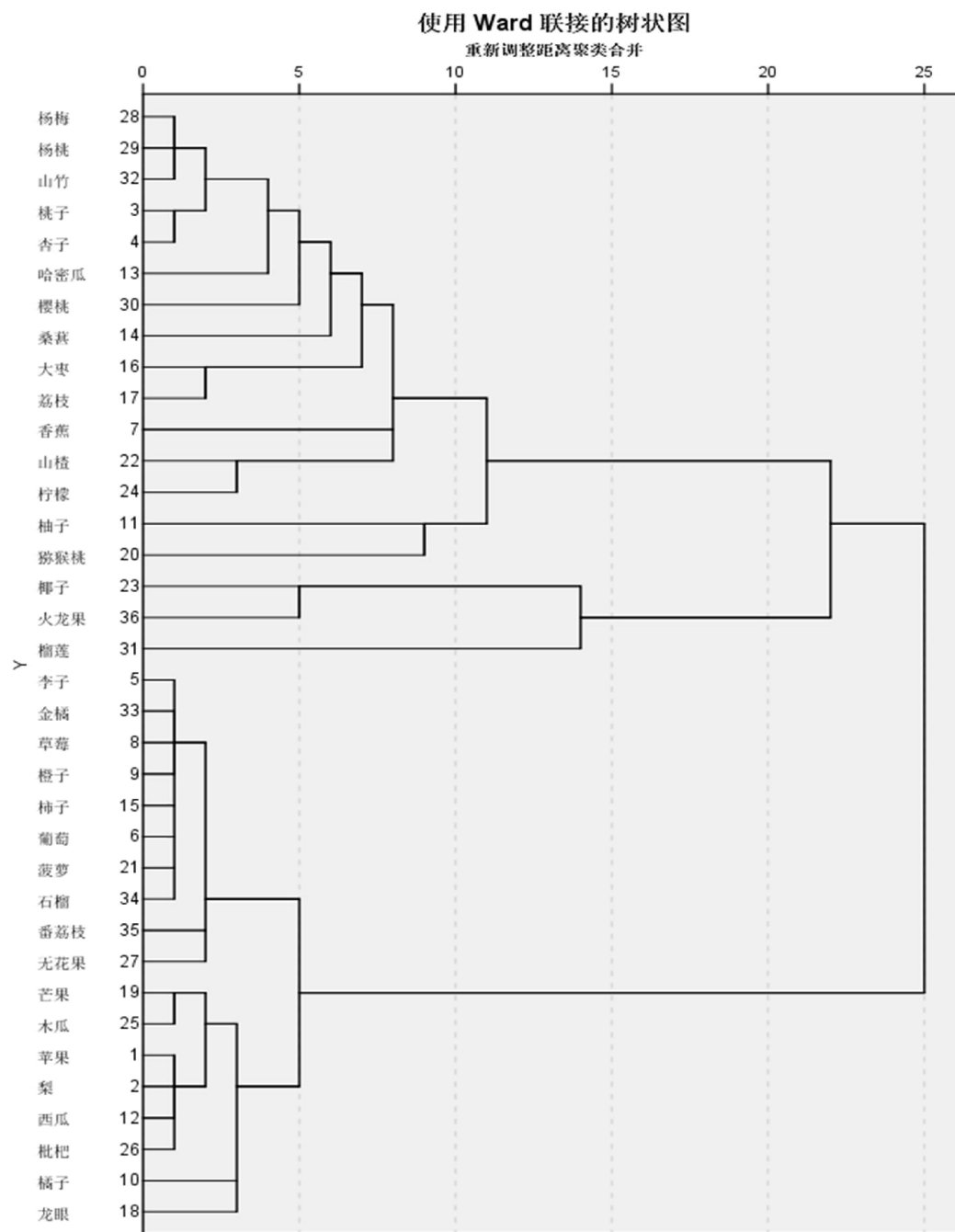


图 19: 45 岁以上水果聚类图

根据聚类树状图，把常见水果营养素对 45 岁以上老年人的效果分为四类：

第一类：柠檬，山楂，香蕉，荔枝，大枣，桑葚，樱桃，哈密瓜，杏子，桃子，山竹，杨梅，杨桃

第二类：猕猴桃，柚子

第三类：榴莲，火龙果，椰子

第四类：龙眼，橘子，枇杷，西瓜，梨，苹果，木瓜，芒果，无花果，番荔枝，石榴，菠萝，葡萄，柿子，橙子，草莓，金橘，李子

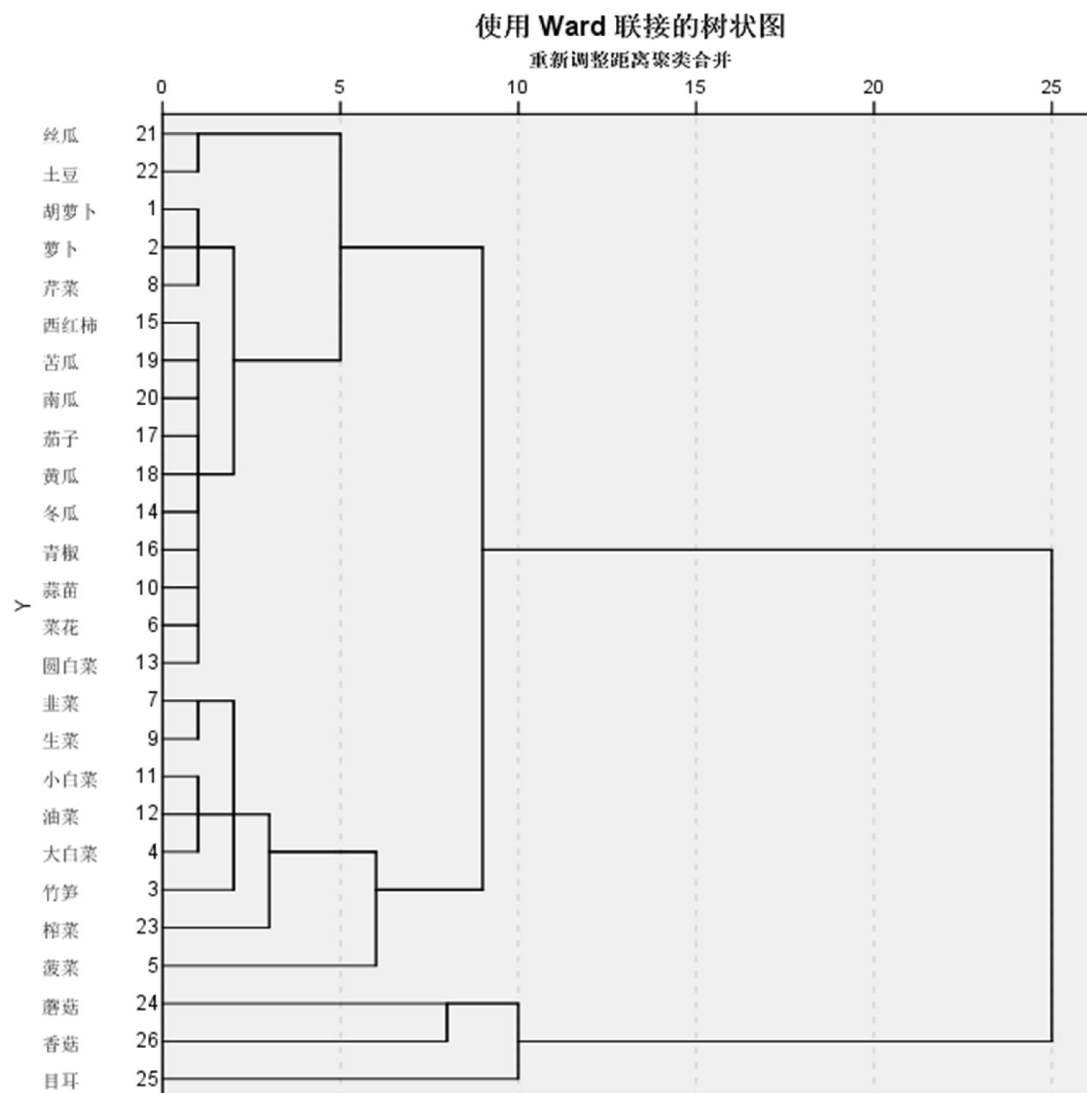


图 20: 45 岁以上蔬菜聚类图

根据聚类树状图，把常见蔬菜营养素对 45 岁以上老年人的效果分为三类：

第一类：圆白菜，菜花，蒜苗，青椒，冬瓜，黄瓜，茄子，南瓜，苦瓜，西红柿，芹菜，萝卜，胡萝卜，土豆，丝瓜

第二类：菠菜，榨菜，竹笋，大白菜，油菜，小白菜，生菜，韭菜

第三类：木耳，香菇，蘑菇

附录 C: MATLAB 源程序代码

```
clc;clear
format long
syms a b;
c=[a b]';
%A=[...];
B=cumsum(A);%原始数据累加
n=length(A);
for i=1:(n-1)
    C(i)=(B(i)+B(i+1))/2;%生成累加矩阵
end
%计算待定参数的值
D=A;D(1)=[];
D=D';
E=[-C;ones(1,n-1)];
c=inv(E*E')*E*D;
c=c';
a=c(1);b=c(2);
%预测后续数据
F=[];F(1)=A(1);
for i=2:(n+10)
    F(i)=(A(1)-b/a)/exp(a*(i-1))+b/a;
end
G=[];G(1)=A(1);
for i=2:(n+9)
    G(i)=F(i)-F(i-1);
end
%t1=2007:2011;
t1=1995:2011;
%t2=2007:2021;
t2=1995:2020;
G=G';
%作对比图
xlswrite('预测值 1.xls',G,1,'a2')
plot(t1,A,'*',t2,G)
```


附录 D：其余果蔬拟合效果图

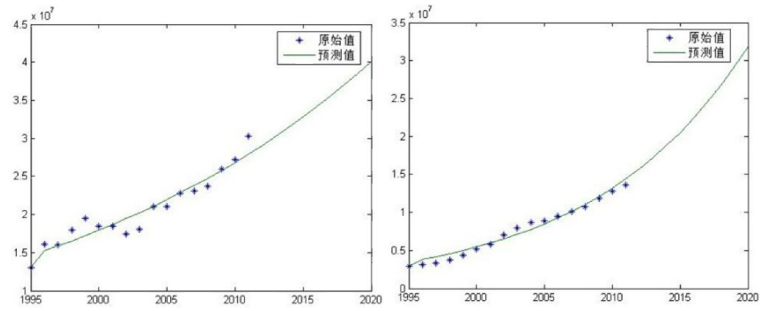


图 21：苹果拟合图

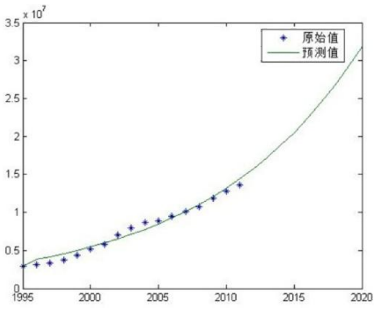


图 22：葡萄拟合图

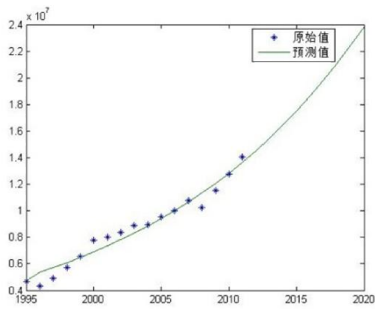


图 23：香蕉拟合图

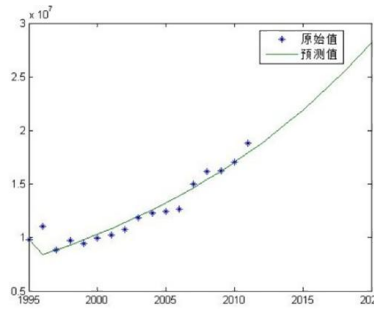


图 24：菜花拟合图

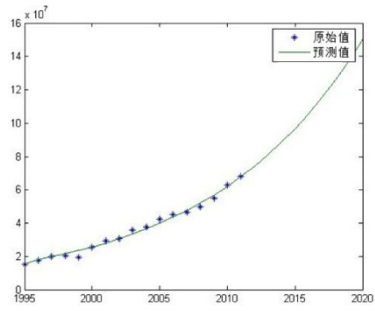


图 25：胡萝卜拟合图

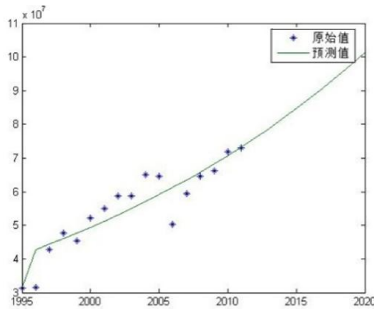


图 26：黄瓜拟合图

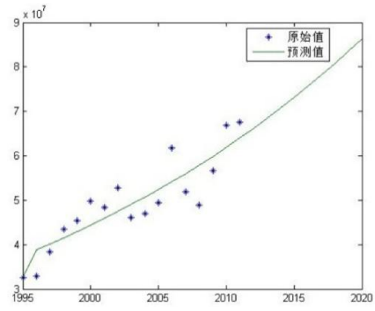


图 27：香菇拟合图

附录 E：1995-2011 年果蔬的的年均产量(来源于国家统计局)

年份	苹果	梨	香蕉	葡萄	蔬菜
1995 年	1400.77	494.24	312.50	174.17	25726.71
1996 年	1704.73	580.66	253.56	188.31	30123.09
1997 年	1721.86	641.48	289.23	203.28	35962.39
1998 年	1948.07	727.55	351.82	235.82	38491.93
1999 年	2080.16	774.23	419.42	270.81	40513.52
2000 年	2043.12	841.24	494.15	328.17	44467.94
2001 年	2001.50	879.61	527.24	367.97	48422.36
2002 年	1924.10	930.94	555.73	447.95	52860.56
2003 年	2110.18	979.84	590.33	517.59	54032.32
2004 年	2367.55	1064.23	605.61	567.53	55064.66
2005 年	2401.11	1132.35	651.81	579.44	56451.49
2006 年	2605.93	1198.61	690.12	627.08	53953.05
2007 年	2785.99	1289.50	779.67	669.68	56452.04
2008 年	2984.66	1353.81	783.47	715.15	59240.35
2009 年	3168.08	1426.30	883.39	794.06	61823.81
2010 年	3326.36	1505.26	956.05	854.90	65099.41
2011 年	3598.48	1579.48	1040.00	906.75	67929.67